

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:47:57 ; Search time 17 seconds
(without alignments)
1721.378 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGPSEAEPAKPKPTS.....RSFWIAIPWACPLGVAALC 562

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2770	92.7	1225	1	DAT1 HUMAN
2	2110	70.6	614	1	DAT1 MOUSE
3	232	7.8	2039	1	PHF3 HUMAN
4	203.5	6.8	656	1	CGBP HUMAN
5	194.5	6.5	660	1	CGBP MOUSE
6	193	6.5	488	1	CYL2 BOVIN
7	189.5	6.3	705	1	TRDN RABIT
8	182.5	6.1	1010	1	HC58 MOUSE
9	174.5	5.8	1012	1	HC58 HUMAN
10	174	5.8	700	1	TRDN CANFA
11	174	5.8	704	1	NP14 RAT
12	172	5.8	5038	1	PCLO MOUSE
13	171.5	5.7	699	1	NP14 HUMAN
14	169	5.7	728	1	TRDN HUMAN
15	167	5.6	1101	1	PHF2 HUMAN
16	164	5.5	1359	1	ATRX CAEL
17	163	5.5	1096	1	PHF2 MOUSE
18	161.5	5.4	1781	1	AKL2 HUMAN
19	161	5.4	2492	1	ATRX HUMAN
20	161	5.4	2492	1	ATRX PANTR
21	161	5.4	2492	1	ATRX PONFY
22	160.5	5.4	2004	1	MY33 HUMAN
23	159.5	5.3	1087	1	NFH MOUSE
24	159	5.3	2476	1	NFH MOUSE
25	158	5.3	755	1	RRBL HUMAN
26	155	5.2	5085	1	PCLO RAT
27	154.5	5.2	1861	1	MAP2 RAT
28	154	5.2	1191	1	CING MOUSE
29	153.5	5.1	6632	1	UN89 CAEL
30	153	5.1	952	1	YK15 CAEL
31	153	5.1	1026	1	NFH HUMAN
32	151.5	5.1	848	1	NFM MOUSE
33	151.5	5.1	1338	1	ACTN MOUSE

ALIGNMENTS

RESULT 1

DAT1 HUMAN
ID DAT1 HUMAN STANDARD; PRT; 1225 AA.
AC Q9BT00; O15043; Q9B0W0; Q9B0W3; Q9H4G6; Q9H4G7; Q9NTU8; Q9N0M8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Death associated transcription factor 1.
GN DAT1 OR KIAA0333.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Placenta;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugeno S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hoshiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saiko K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEBO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

[2]
SEQUENCE FROM N.A.
MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagge C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Humble B., Hunt A.R., Hunt S.E., Jekesch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharshaino M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.D., Soderlund C., Steward C.W., Sulston J.E.,
RA Swann R.M., Symanore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871 (2001).

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC TISSUE=Colon, Kidney, and Muscle;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.R.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., McQuellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.G., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton S., Kertanen M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Elakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalus D.B.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 235-1225 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Korani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL GNA Res. 4:141-150(1997)
CC -!- FUNCTION= putative transcription factor, weakly pro-apoptotic when
CC overexpressed (By similarity).
CC -!- SUBCELLULAR LOCATION= Cytoplasmic; translocates to the nucleus
CC after pro-apoptotic stimuli (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC Name=2;
CC IsoId=Q9BTC0-1; Sequence=Displayed;
CC IsoId=Q9BTC0-2; Sequence=VSP_007208, VSP_007209, VSP_007210;
CC Note= No experimental confirmation available;
CC Name=3;
CC IsoId=Q9BTC0-3; Sequence=VSP_007208, VSP_007211, VSP_007212;
CC Note= No experimental confirmation available;
CC -!- TISSUE SPECIFICITY= Ubiquitous.
CC -!- SIMILARITY= Contains 1 PBD-type zinc finger.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK002127; BA932094.1; -
CC EMBL; AL035669; CAC12755.1; ALT_SEQ.
CC EMBL; AL035669; CAC12755.1; -
CC EMBL; AL035669; CAC28883.1; -
CC EMBL; AL117379; CAB90275.1; -
CC EMBL; BC000770; AAH00770.1; -
CC EMBL; BC004237; AAH04237.1; -
CC EMBL; BC014489; AAH14489.1; -
CC EMBL; AB002331; BA920791.1; -
CC Genew; HGNC:2680; DATF1.
CC
CC MIM; 604140; -
CC InterPro; IPR003618; TFS2_centre.
CC InterPro; IPR001965; Znf_PBD.

DR Pfam; PF00628; PHD; 1.
DR SMART; SM00349; PHD; 1.
DR SMART; SM00510; TFS2M; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS030016; ZF_PHD_2; 1.
KW Apoptosis; Nuclear protein; Zinc-finger; Alternative splicing.
FT DOMAIN 165 173 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN FING 185 193 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN FING 268 322 PHD-TYPE.
FT VARSPPLIC 388 423 Missing (in isoform 2 and isoform 3).
FT VARSPPLIC 566 580 /FTId=VSP_007208. /FTId=VSP_007209. (in isoform 2).
FT VARSPPLIC 581 1225 Missing (in isoform 2).
FT VARSPPLIC 566 598 /FTId=VSP_007210. /FTId=VSP_007211. (in isoform 3).
FT VARSPPLIC 599 1225 Missing (in isoform 3).
FT CONFLICT 252 252 G -> E (IN REF. 1).
FT CONFLICT 580 580 T -> M (IN REF. 2).
FT CONFLICT 592 592 T -> A (IN REF. 2).
SQ SEQUENCE 1225 AA; 133023 MW; 595071E0D3CC74A2 CRC64;

Query Match 92.7%; Score 2770; DB 1; Length 1225;
Best Local Similarity 93.6%; Pred. No. 5.5e-141;
Matches 529; Conservative 0; Mismatches 0; Indels 36; Gaps 1;

Qy 1 MDDKGDPSNEAPKAIKPTSKFRKTWGRPTTIAREGGAGDAEADPLEPPPOOGLS 60
Db 1 MDDKGDPSNEAPKAIKPTSKFRKTWGRPTTIAREGGAGDAEADPLEPPPOOGLS 60

Qy 61 LRSGRQPKTERVEQPLTIARRGRSMVPSLEDSCPTSCPATDAETASEGVESASE 120
Db 61 LRSGRQPKTERVEQPLTIARRGRSMVPSLEDSCPTSCPATDAETASEGVESASE 120

Qy 121 TRSGQASATAVERPASSEKVGDDHDDTSDSDGLTKELQNLRLRREOEPTERP 180
Db 121 TRSGQASATAVERPASSEKVGDDHDDTSDSDGLTKELQNLRLRREOEPTERP 180

Qy 181 LKGIOSRLRKRREEGPAETVGSASDTVGVLPKQEPENDQGVVSOAGKDRESKLE 240
Db 181 LKGIOSRLRKRREEGPAETVGSASDTVGVLPKQEPENDQGVVSOAGKDRESKLE 240

Qy 241 KAAQDIKDREPGDGRPKPECEGDPNLYCICRQPNRPFMICCDRCBWFHGDVIGIS 300
Db 241 KAAQDIKDREPGDGRPKPECEGDPNLYCICRQPNRPFMICCDRCBWFHGDVIGIS 300

Qy 301 EARGIALERNGEDYICENCTILOVDETHSETAQQAQKWRPGDADGTDCTICTIQKS 360
Db 301 EARGIALERNGEDYICENCTILOVDETHSETAQQAQKWRPGDADGTDCTICTIQKS 360

Qy 361 SEDQIGIKRIEKAANPSGKKLIFQP----- 387
Db 361 SEDQIGIKRIEKAANPSGKKLIFQP----- 387

Qy 388 ----VIEAPGASKICGPOCCHVAQDSDVYCSNDCTILKHAATMKFLSGKQKPKPKMK 444
Db 421 SKRVIEAPGASKICGPOCCHVAQDSDVYCSNDCTILKHAATMKFLSGKQKPKPKMK 480

Qy 445 MKPEKPSLPKCGAQAQIKISSVHKRPAPKPKETTVKKAIVVVPARFSEALKEACESTPS 504
Db 481 MKPEKPSLPKCGAQAQIKISSVHKRPAPKPKETTVKKAIVVVPARFSEALKEACESTPS 540

Qy 505 WADSHNTNAVPKPKTAAPSPSLLYK 529
Db 541 WADSHNTNAVPKPKTAAPSPSLLYK 565

RESULT 2

DAT1_MOUSE STANDARD; PRT; 614 AA.
AC Q89B9; Q89B2; Q9WV00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Death associated transcription factor 1 (Death inducer-obliterators-1) (DIO-1).
DE DAT1 OR DIO1 OR DIO1.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP TISSUE=Pre-B cell;
RC MEDLINE=93324176; PubMed=10393935;
RX Garcia-Domingo D., Leonardo B., Grandien A., Martinez P., Albar J.P., Izpisua-Belmonte J.-C., Martinez-A C.;
RA "DIO-1 is a novel gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997 (1999).
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik J., Gough J., Grimmond S., Gutierrez S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konojaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wyshaw-Boris A., Yangisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
CC -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when overexpressed.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus after pro-apoptotic stimuli.
CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
CC -!- INDUCTION: Upregulated during apoptosis.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ238332; CAB48401.1; -;
CC EMBL; AK042474; BAC31270.1; -;

DR EMBL; AK044919; BAC32141.1; -;
DR MGD; MGI:1344352; Dicol.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0006915; P:apoptosis; IDA.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
KW Apoptosis; Nuclear protein; Zinc-finger.
FT DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 182 190 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN FING 265 319 PHD-TYPE.
FT CONFLICT 45 45 V -> A (IN REF. 1).
FT CONFLICT 331 331 D -> N (IN REF. 1).
FT CONFLICT 333 333 I -> V (IN REF. 1).
SQ SEQUENCE 614 AA; 67409 MW; 15ACCS4BAEF3539 CRC64;
Query Match 70.6%; Score 2110; DB 1; Length 614;
Best Local Similarity 75.2%; Pred. No. 4.9e-106;
Matches 421; Conservative 30; Mismatches 95; Indels 14; Gaps 7;
QY 1 MDDKGPSNEBAPKAIPKTSKFRKTWGRRTIARRGRSRMPVLSDESGEPTSCPATDAETASEGSVASE 60
DB 1 MDDKGLSNEBAPKAIPKTSKFRKTWGRRTIARRGRSRMPVLSDESGEPTSCPATDAETASEGSVASE 58
QY 61 LRSGEQPQKTRERVOFLTIARRGRSRMPVLSDESGEPTSCPATDAETASEGSVASE 120
DB 59 LRSGEQPQKTRERVOFLTIARRGRSRMPVLSDESGEPTSCPATDAETASEGSVASE 118
QY 121 TRSGPOSATAVKRPASSEKVKGGDDHDDTSDSDGLTLKELQNRIRKRKEQTERP 180
DB 119 IRSGPVSLSLG-KEHPASSEKARGGEEDTSDSDGLTLKELQNRIRKRKEQTERP 177
QY 181 LKGIQRLKRREREGPASTVGEASDVGVLP-SKQEPENDQGVVSOAGKDDRESKLE 239
DB 178 LRSGQNRKRREREDSALT-GEVIGSREQRPLCKQEPASQGVPSQSEYDDIENQLE 236
QY 240 GKAAQIDKDEPDLGRPKPECEGVDPNLYCICRPHNNRPMICDRCEWPHGDCVGI 299
DB 237 GRATQNTENPREAGKPKPECEVDPNLYCICRPHNNRPMICDRCEWPHGDCVGI 296
QY 300 SRAGELLERNGEDYICPNTILQVDETSTADQCEAKWPGDADGTDCTSIGTIEOK 359
DB 297 SRAGELLERNGEDYICPNTILQVDETSTADQCEAKWPGDADGTDCTSIGTIEOK 356
QY 360 SSEDQIKRIEKAANPSGKKLKIQQPVIEAPGASKICPGCGCHVAQEDSVYCSNDCIL 419
DB 357 SGEDQIKRIEKAANPSGKKLKIQQPVIEAPGAPKICPGCGSVAPQVSVYCSNDCIL 416
QY 420 KRAAATMKTLSGKEQKPKPKKKKKPKPSLPKGAQAGIKLSSVHKRPAPKEKETT 479
DB 417 KRAAATMKTLSGKEQKPKPKKKKKPKPSLPKGSVQVGIKISSVHKRLASEKKNPV 476
QY 480 KKAVVVPASSEALGKEACESSTPSWASDHNNVAVKPEKTAAP---SPSLLYKCMYHLGV 536
DB 477 KK-VMLASSETSGKEACESSTPSWASDHNNVAVKPEKPEKPTALSPILLSACTYHPKA 535
QY 537 GLLDPSRSFWIAIPWACPL 556
DB 536 GFPGPSHLL-----GGCLGL 550
RESULT 3
PHF3_HUMAN STANDARD; PRT; 2039 AA.
ID PHF3_HUMAN
AC Q92576; Q9NQ16; Q9UI45;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PHD finger protein 3.
OS Homo sapiens (Human).

1	MEG----	DGSDPEPDAGEDS--KSENGENAPIYCI	CRKPDINC	FWICGDCN	NEWFHGD	1
298	GISEANGRLLENGEDYICPNC	-----TILQVQ-----	DETHSETADQOEAKWR	341		
56	RITERMAKAI-----	REWYCREKEDPKLEIRYRHKTSRERD	GNERSDESSEPRDGGGRKR	111		
342	P-CDADGTCTSIGTITEQSSSDQIGKRIEK	AANPSGKKLKFQVIEAP	-----	392		
112	PVDPD-----	-----LQVAGSGTGVMALANGS	SPHKSSP-----	QDLYVTFQHQHQ	158	
393	-----GASKCIGPGCHVAQPSD	SVYCSNDICILKEAAATMKPLSGKEQKPKPKEMK	-----	444		
159	QQQIKESARMC-----	GBCEACRR-----	TEDC-----	GHCDPFRDMKKFGGPNKIRQKRLRQC	208	
445	-----MKPEP-----	SILPKCGAAGIKISSVHKPEAP	PEKKTETTKVAVVVPAR	488		
209	QLEARESYPKSSSPVTFSES	LP-----	PRPLPTQQQ-----	POP	247	
489	SEALGK-----	EAACSSSTPSWASDHNNAVK-----	PRKTAAPSP	524		
248	SQKLGRIRDEGAVASST-----	VKEPPEATATPEP	278			
RESULT 5						
ID	CGBP MOUSE	STANDARD;	PRF;	660 AA.		
AC	QSCW7;					
DT	28-FEB-2003 (Rel. 41, Created)					
DT	28-FEB-2003 (Rel. 41, Last sequence update)					
DE	CPG binding protein (protein containing PHD finger and CXXC domain 1).					
GN	CGBP OR PCCX1.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_Taxid=10090;					
XP	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;					
FX	MEDLINE=21085660; PubMed=11217851;					
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,					
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,					
RA	Arakawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,					
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,					
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,					
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,					
RA	Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,					
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,					
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,					
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,					
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,					
RA	Gustinch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,					
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,					
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,					
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,					
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,					
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,					
RA	Hayashizaki Y.					
RT	Functional annotation of a full-length mouse cDNA collection."					
RL	Nature 409:685-690(2001).					
CC	FUNCTION: Transcriptional activator that exhibits a unique DNA					
CC	binding specificity for [AC]CpG[AC] unmethylated CpG motifs (By					
CC	similarity).					
CC	SUBCELLULAR LOCATION: Nuclear (By similarity).					
CC	SIMILARITY: Contains 1 PHD-type zinc finger.					
CC	SIMILARITY: Contains 1 CXXC-type zinc finger.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)					
CC	or send an email to license@isb-sib.ch					
EMBL	AF149758; AAF37799.1; -					
EMBL	AB031069; BAA96307.1; -					
EMBL	AL136822; CAB66796.1; -					
EMBL	BC014940; AAH14940.1; -					

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK010337; BAB26862.1; -.
DR MGD; MG1:1921572; Cdbp.
DR GO; GO:0016343; C:nuclear matrix; IDA.
DR InterPro; IPR002857; Znf_CXKC.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; ZF-CXKC; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
KW Metal-binding; Coiled coil; Nuclear protein.
FT ZN_FING 28 76
FT ZN_FING 165 212
FT DOMAIN 260 321
FT DOMAIN 325 363
FT DOMAIN 426 479
FT DOMAIN 136 161
FT DOMAIN 166 225
FT DNA_BIND 166 225
SQ SEQUENCE 660 AA; 76166 MW; 9D92CB6DA82069EA CRC64;

Query Match 6.5%; Score 194.5; DB 1; Length 660;
Best Local Similarity 22.9%; Pred. No. 0.0012;
Matches 78; Conservative 46; Mismatches 104; Indels 113; Gaps 18;

QY 238 LEGKAAQDIKEEPPDGLGRPECEGYDPMALYICICROPHNRPVICCDRCZEFHGDV 297
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1 MEG-----DGSLEPPDAG--DSSKSENGENAPIYICICRXPDPNCFMIGDCNENFPHGDCI 55
QY 298 GISEARGLLERNGDYICPNC-----TILQVQ-----DETHSETADQOEAQR 341
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
56 RITERKAKAI-----REWCYRECKDKLEIRYHKKCRERDGSERAGEPRDEGGGRK 111
QY 342 PGADGTDCTSIGTIEKSSDEQGTGRIEK--AANPSGKKKLIKIPQVIEAP----- 392
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
112 P-----ASDPELQRRAGSGTGVGAMLRGASAPHSKSP-----QPLVATPSQHHQOQ 159
QY 393 -----GASKCTGPGCCHVAQPDVSVCNSDILKHAATMKFLSSGKQKPKPKEM 443
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
160 QQQQQQIKRSARMC---GECEACRR-----TEDC--GHCDFCRDMMKFGGNKINQKRL 209
QY 444 K-----MKPEKPS--LPKCGAQAQGIKISSVHKRPAPKPKETTKVAVV 485
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
210 RQCLRARESYKVPSSLSPTVPSALPR-----PRRPPTQOQ----- 248
QY 486 PARSEALGKEACESTPSWASHNYNAVK--PEKTAAPSP 524
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
249 PQOSOKLGRIREDEGTVLVS-----SVVKEPPEATATPEP 282

RESULT 6
CYL2_BOVIN STANDARD; PRT; 488 AA.
AC Q28092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cylicin II (Multiple-band polypeptide II).
GN CYL2 OR CYL2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95255491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
heads: the identification and characterization of cylicin II.";
```

```

Exp. Cell Res. 218:174-182(1995).
[2]
RN PARTIAL SEQUENCE.
RP MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
the sperm head cytoskeleton."
J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: Possible architectural role during spermatogenesis. May
be involved in spermatid differentiation.
CC -!- SUBCELLULAR LOCATION: Calyx; sperm head cytoskeletal structure.
CC -!- TISSUE SPECIFICITY: Testis.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46789; CNA86753.1; -.
DR PIR; I46014; I46014.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 487
FT DOMAIN 189 407
FT REPEAT 189 212
FT REPEAT 213 239
FT REPEAT 240 267
FT REPEAT 268 295
FT REPEAT 296 323
FT REPEAT 324 351
FT REPEAT 352 379
FT REPEAT 380 407
SQ SEQUENCE 488 AA; 53562 MW; 5F3107CB2D019C5A CRC64;

Query Match 6.5%; Score 193; DB 1; Length 488;
Best Local Similarity 20.3%; Pred. No. 0.0011;
Matches 117; Conservative 73; Mismatches 204; Indels 182; Gaps 19;

QY 18 PTKSEFKTGCFRNTTIKREGAGDAEADPLEPPPPQQLGLSLRRSGRQPRTERVEQF 77
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
19 PVSLSKKSN-----QQHTALVFKPRPKR----- 46
QY 78 LTIARRGRSRMPVSLDSGEPTSCPATDAF-----TASEGSVESAS 119
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
47 -----RSKPSLLQEN-----TSPKYDAEKLGRDRQPLMHRSLMRISERPVSYLAA 94
QY 120 ETR-----SGPOSASTAVKERPASSEKVKGGDDHDDTSDSDGLTLKELQNLRRKREOE 175
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
95 RSRHPQKETPPSQBDAAQAAKPSKVKKSKEDKDKSDSEASISVSKPKPKLKAKEEK 154
QY 176 PTER-----PLKGTQS-----RLKKRREREGPAETVGSASDT 208
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
155 PDEKKDLKKERKDSKKGKESATESEDEDKAGAEKAGKDKKSGKKGKETSPTS--GSEKGA 213
QY 209 VEGVLPKQPEPE-----NDQGVVSQAGKDDRESK-----LEGKAAQDIKEE 250
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
214 KDSKSKKDSKKGKESATESESEGKDAKDDKDKKSGKKGKESATESESEGKDAKDDK 273
QY 251 PGDLGRPK-----PECEGYDPMALYICICROPHNRPVICCDRCZEFHGDVCGISEARGR 305
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
274 KKKKSGKKGKESATESESEGKDAK-----DDKKGK 304
QY 306 LLENGEDYICPNCTILQVDETHSETADQOEAQRPGDADGTCTSIGTIEKSSDEQ 365
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
305 KSGKKGK-----SATESEGKDAKDDKDKKSGKKGK-----KESATESEG 347
QY 366 IKGRIEKAANPSGKKKLIKIPQVIEAPGASGICIGPCCHVAQPDVSVCNSDILKHAAT 425
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
348 EKGDAK--DDKKGK-----GSKK--GKESDSKAEQDKGDAKD--DKDKKGS 391
QY 426 MKFTSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 485
```


Db 336 KKSKSEKEDIDMKKK-----PFGKSPDKPTGTVKVT-----QAATK 372
 Qy 427 ---KFLSSGKEKXP---PKKPKKKPKPKSPKPKCGA-----QAGIKI-----463
 Db 373 KDEKEDSKKAKPAEOPFGKQKQKKEHEBPAKSTKKEHAAPSEKQAKAKIERKERS 432
 Qy 464 SSVHRRPAPEKKTITTKKAVVVPASSEALGKAAC-----ESSTPPSWASDHNTNA 513
 Db 433 AASTTKKAVPAKKEKTKTVQETKKEKPKGKISSVLKDKELTKKEVKVPASLKEKGSSET 492
 Qy 514 VKPEKTAAPSPSL 526
 Db 493 KDKKETSXPEQOI 505

RESULT 8

HC58 MOUSE
 ID HC58 MOUSE STANDARD; PRT; 1010 AA.
 AC Q8BLG0; Q8BMA2; Q8BYR4; Q921N1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hepatocellular carcinoma-associated antigen 58 homolog.
 GN HC58.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RP STRAIN=C57BL/6J; TISSUE=Embryo, and Hypothalamus;
 RX MEDLINE=22354683; PubMed=12456851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Baldarrelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Naita K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszynski B., Yang Y., Yang I., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs.
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE OF 621-1010 FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cavaletto T.L., Scheetz T.B.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.E.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krywinski M.I., Skalska U., Smallus D.E.,
 RA Smercher A., Schein J.E., Jones S.J.M., Matra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Possible transcription factor (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8BLG0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8BLG0-2; Sequence=VSP_007762;
 CC Note=Sequence incomplete. No experimental confirmation
 CC available;
 CC -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AK033017; BAC28129.1; -
 CC EMBL; AK038573; BAC30050.1; -
 CC EMBL; AK045309; BAC32304.1; -
 CC EMBL; BC011337; AAH11337.1; -
 CC MGD; MGI:2444148; 682040202ORik.
 CC InterPro; IPR002999; Tudor.
 CC InterPro; IPR007087; Znf_C2H2.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF00628; PHD; 1.
 CC Pfam; PF00096; Zf-C2H2; 1.
 CC SMART; SM00249; PHD; 1.
 CC SMART; SM00333; Tudor; 2.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.
 CC PROSITE; PS01359; ZF_PHD_1; 1.
 CC PROSITE; PS00016; ZF_PHD_2; FALSE NEG.
 CC Transcription regulation; Nuclear protein; DNA-binding; Metal-binding;
 CC Zinc; Zinc-finger; Alternative splicing.
 CC DNR_BIND 257 269 A.T.HOOK (BY SIMILARITY).
 CC ZN_FING 455 485 C2H2-TYPE.
 CC ZN_FING 657 703 PHD-TYPE.
 CC FT DOMAIN 162 259 LYS-RICH.
 CC FT DOMAIN 526 555 LYS-RICH.
 CC FT DOMAIN 543 555 POLY-LYS.
 CC FT VARSPLIC 1 28
 CC WTKEPPNPGISFFVGAQLRDLKKN -> GAARTVLLS
 CC VGLERRSRSGAVR (in isoform 2).
 CC /FTID=VSP_007762.
 CC FT CONFLICT 817 817 A -> T (IN REF. 2).
 CC SEQUENCE 1010 AA; 115279 MW; 2BC32811A520342D CRC64;
 SQ

Query Match 6.1%; Score 182.5; DB 1; Length 1010;
 Best Local Similarity 20.3%; Pred. No. 0.0085;
 Matches 143; Conservative 78; Mismatches 252; Indels 231; Gaps 31;
 Qy 3 DKGDPSENEAPKAIKPTSKFRTKWTGFRFTTKRSE-----38
 Db 304 DKNSP-QQSKKRKSENSDKILSRSSRLSTNGTRELIDPDPSIVPDLVHTVTNPLPKDS 362

childhood medulloblastoma cDNA libraries.";
Int. J. Cancer 106:244-251(2003).
-!- FUNCTION: Possible transcription factor.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q9BVI0-1; Sequence=Displayed;
Name=2;
IsoId=Q9BVI0-2; Sequence=VSP_007760, VSP_007761;
Note=Ref.1 isoform 2 is in conflict in position(s):
237-239:KTR->KTS;
-!- TISSUE SPECIFICITY: Expressed in heart, kidney, liver, lung, pancreas, placenta, spleen and testis. Not expressed in brain, skeletal muscle, colon, ovary, prostate, small intestine and thymus. Expressed in colon and ovary cancer cell line, whereas it is not expressed in the respective normal tissues.
-!- DISEASE: Antibodies against C20orf104 are present in sera from patients with hepatocellular carcinoma, glioblastoma and childhood medulloblastoma who developed different autoantibodies. May serve as diagnostic marker in glioblastoma.
-!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
-!- SIMILARITY: Contains 1 PHD-type zinc finger.
-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- CAUTION: Ref.1 sequence differs from that shown due to a frameshift in position 260.
-!- CAUTION: Ref.2 (AAK19748) sequence differs from that shown due to frame shifts in positions 635, 653 and 655.
-!- CAUTION: Ref.4 sequence differs from that shown due to a chimeric cDNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AF220416; AAF34184.1; ALT_FRAME.
EMBL; AY027523; AAK13046.1; -.
EMBL; AF348207; AAK19748.1; ALT_FRAME.
EMBL; AL078461; CAC09384.1; -.
EMBL; AL109965; CAC09389.1; -.
EMBL; AF258787; AAG49888.1; -.
EMBL; BC001198; AAK01198.1; ALT_SEQ.
EMBL; BC048210; AAK48210.1; -.
Genew; HGNC:16098; C20orf104.
InterPro; IPR002999; Tudor.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR001965; Znf_PHD.
Pfam; PF00628; PHD; 1.
Pfam; PF00096; Zf-C2H2; 1.
SMART; SM00249; PHD; 1.
SMART; SM00333; Tudor; 2.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.
PROSITE; PS01359; ZF_PHD_1; 1.
PROSITE; PS00016; ZF_PHD_2; FALSE NEG.
Transcription regulation; Nuclear protein; DNA-binding; Metal-binding; Zinc; Zinc-finger; Alternative splicing.
FT DNA BIND 257 269 A.T HOOK.
FT ZN FING 452 477 C2H2-TYPE.
FT ZN FING 654 700 PHD-TYPE.
FT DOMAIN 162 221 LYS-RICH.
FT DOMAIN 523 552 LYS-RICH.
FT VARSPLIC 237 273 VDKPNDIVKSPQENLRPKRGPPSIAPTVDS ->
KTRQPFHSSVCCGFLSNPATNIGTEKENIKEM (in isoform 2).
FT FT /FTID=VSP_007760.
FT VARSPLIC 274 1012 Missing (in isoform 2).
FT FT /FTID=VSP_007761.
FT CONFLICT 221 222 KE -> QG (IN REF. 5).

FT CONFLICT 226 226 E -> G (IN REF. 5).
FT CONFLICT 516 516 S -> F (IN REF. 5).
SQ SEQUENCE 1012 AA; 115385 MW; 1CDBADC23D007503 CRC64;
Query Match 5.8%; Score 174.5; DB 1; Length 1012;
Best local Similarity 19.3%; Pred. No. 0.023;
Matches 156; Conservative 72; Mismatches 269; Indels 285; Gaps 32;
QY 3 DKGDPNREAPKAKPTSKFRXTWGRFRRTIAKREGAGDAEADPLEP-----PPPQQQL 57
DB 204 EKGVSEKSLPKNEKEDKENISEN-----DREVSQDAQVDK-KPENDIVKSPQENL 253
QY 58 GLSLRSGROPK-----RTVEQFLITARR----- 84
DB 254 REPKRGRGPPSIAPTVDSNSQTLPITLELRKIKSGCEVPLKRPRLDKNSSQBSKSK 313
QY 85 -----GRSRMPVSLDSGE-----PTSCPATDAETAEGSVESA 118
DB 314 NYSNTDKLSRRRSFSLSTNGTHEILDPLVVDLVDTPDLQUTLSSTKESEGQKSA 373
QY 119 SPT-----RSG-----POSATAKVERPASS-----EKYGGDDHD 149
DB 374 LEAQVSSALTCHSGFGSGAAGLELNCPSMGENTMTKTEPTSLVLEQLSTVEVTTFK 433
QY 150 DTSDDSDGUTLKEIQLNRLRK-----REQEPTERPLK-G 183
DB 434 KTDGSSNAPAVDLDHFKCVKVDCLKPPRKAKLLHYEMKYTHGMEKSLSESPGKRH 493
QY 184 IQSR-----LRKGRREGP 197
DB 494 VQTRGSPASDKPSQETLTRKRVSSASPTTNDKKNKEKFKFVRVVKPKKKKKKKTKP 553
QY 198 AETVGSASDTVGVLPSK-----OBPENDQGVVSQAGKDD- 233
DB 554 ECPSEISITSEPPPKAPATRCSSHKPGVHMSPOLHGPESHGKVKVXALBEDNL 613
QY 234 RESKLEGAQADINDEPPGDLG---RPKPCEG---YDPNALYICIQPHNNRPMICDR 287
DB 614 SSSSSSFLMSD--DEYGDQVDVTNPDDELGDGDDRYDFEVVRCICEVQENDFMIOCEE 671
QY 288 CEWFHGDGCVGISSEARGILERN-GDYICPNTILQVQDTHSETADQAEAKWRPGDAD 346
DB 672 CQWQHGVCMG-----LLERNVPEKITCYVC-----QDPGQRPFGKY 709
QY 347 GTD-----CTSIGTIEKSSSEDQIK-----GRIEKAAN--PSGKKKLIKIPVP 388
DB 710 WYDKWLSRGHMGHGLAFLEENYSHONAKKIVATHQLIGDVQVIEVLHGLQKMSILQS- 768
QY 389 IBAPGASKICPGCCHVAQPDVYCS---NDCILKHAATWKFLSCGKCKPKKMKM 445
DB 769 RHPDPLPWCOPWKQSHSGRSHFRNIPVTDTRSKERPSYRIL-NGAVEKPREL----- 822
QY 446 KPEKPSLPKCAQAGIKISSVHKRPAPKEKETTIVKKAIVVVPVPAVSEALG-----KEAAACE 499
DB 823 -----ALPLPRSVESYITSEHCYQKPRAYVPAVBOKLWVETRGSALDDAVNPLHENGDD 877
QY 500 SSTP--SWASDHNYNA-----VKP-----EKTAAPSILLYKMHVGLGVLLDPSRS 544
DB 878 SLSPGLGWFLDQDRSGKSDPKFGSPKVKYVSKALPPEAPARKLLDRGEGGLLS-SQH 936
QY 545 FW 546
DB 937 QW 938
RESULT 10
TRDN CANPA STANDARD; PRT; 700 AA.
ID TRDN CANPA
AC P82179;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triadin.

Biochem. Biophys. Res. Commun. 230:370-375(1997).

[5]
CHARACTERIZATION.

RA MEDLINE=200336810; PubMed=10567578;
Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;
human Nopp140, which interacts with RNA polymerase I: implications
for RNA gene transcription and nucleolar structural organization.";
Mol. Cell. Biol. 19:8536-8546(1999).

-1- FUNCTION: Related to nucleogenesis, may play a role in the
maintenance of the fundamental structure of the fibrillar center
and dense fibrillar component in the nucleolus. It has intrinsic
GTPase and ATPase activities. May play an important role in
transcription catalyzed by RNA polymerase I.

-1- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPA194)
and with casein kinase-II.

-1- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the
cytoplasm. At telophase it begins to assemble into granular-like
pre-nucleolar bodies which are subsequently relocated to nucleoli
at the early G1-phase.

-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Alpha;
IsoId=Q14978-1; Sequence=Displayed;
Name=Beta;
IsoId=Q14978-2; Sequence=VSP_004338;

-1- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
cycles on Ck2 and PKC sites. There is evidence suggesting that
CDC2 kinase phosphorylates p130 at the M-phase.

-1- SIMILARITY: Contains 1 Lish domain.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z34289; CAA84063.1; -;
EMBL; D21262; BAA04803.1; -;
PIR; I38073; I38073.
Genew; HGNC:15608; NOLC1.
GK; Q14978; -;
MIM; 602394; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005730; C:nucleolus; TAS.
GO; GO:0007049; P:cell cycle; TAS.
GO; GO:0007067; P:mitosis; TAS.
GO; GO:0006364; P:rRNA processing; TAS.
InterPro; IPR006594; Lish.
InterPro; IPR007718; SRP40_C.
Pfam; PF05022; SRP40_C; 1.
SMART; SM00667; Lish; 1.
PROSITE; PS50896; LISH; 1.
KW Nuclear protein; phosphorylation; Repeat; GTP-binding; ATP-binding;
KW Alternative splicing.
KW DOMAIN 10 42
FT DOMAIN 84 566
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 470 481
FT REPEAT 519 529
FT REPEAT 555 566
FT DOMAIN 68 82
FT DOMAIN 204 382
FT DOMAIN 384 587
FT REPEAT 84 95
FT REPEAT 125 136
FT REPEAT 167 178
FT REPEAT 221 232
FT REPEAT 264 275
FT REPEAT 325 336
FT REPEAT 363 375
FT REPEAT 425 436
FT REPEAT 47

FT	DOMAIN	601	617	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL). PHOSPHORYLATION (BY CK2) (BY SIMILARITY). K -> KWITTSVRAB (in isoform Beta). /FTid=VSP_004338.
FT	MOD RES	563		
FT	VARSP LIC	241	241	
FT	CONFLICT	3	3	D -> A (IN REF. 2).
FT	CONFLICT	133	133	R -> S (IN REF. 2).
FT	CONFLICT	291	292	YA -> SV (IN REF. 2).
FT	CONFLICT	456	456	S -> P (IN REF. 2).
FT	SEQUENCE	699 AA;	73720 MW;	DFD4AD94EDF659FB CRC64;
Query Match		5.7%;	Score 171.5;	DB 1; Length 699;
Best Local Similarity		19.5%;	Pred. No. 0.022;	
Matches 108;	Conservative	84;	Mismatches 218;	Indels 145; Gaps 19;
QY	2	DDKGDPSNEEAPKAIKPTSKTEFKTWGFRRTTAKREGAGDAEAD-PLRPPPPQQQLGLS	60	
Db	136	DBEDQKQPVQGVKVPQAKAAP-----PKAKSSDSDSSEBEPENQ-----	183	
QY	61	LRRSGRPKTEERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASGVSSEASE	120	
Db	184	-----KPKIT-----PVTVAQTAPKPKAPAPKPIANGKAASSS	219	
QY	121	TRSGPOSASTAVKERPASSBK-----VKGEDDHDITSDSDSLGLTLKELQNR	167	
Db	220	SSSSSSSSDSSEBKAATPKTVPKKQVAKAPVKAATTPTRKSSSESDSSDEBEBQK	279	
QY	168	LRRKREOEP-----TERPLKGIORLKKRREGEPAFTVSGEADTVEGVLPKQEP	219	
Db	280	KPMKNKPGPYSAFPSPAPPKPKSLGTPPKKAVEKQPVSESSDSDESSESEBCKP	339	
QY	220	ENDQGVVSQAGDDRESKLEGAQDIKDEPGDLGRPKPCEGYDNAL-----	269	
Db	340	PT-KAVYSKATTKPPAPKAAEISSDSSESDSDEBAPSKPATGTTKNSNKPATVTKSP	398	
QY	270	-----YCICRPHNNRWFIMCCDRCEWPHGVCVGISSEARGLLEHNGEDYICPNCTILQVQ	325	
Db	399	AVKFAAPAKP-----VGGGQ---KLLTRKAD-----SSS	425	
QY	326	DETHSETADQOEAK-----WRPGDAGTDCISI-----GTIEQKSESDGIGRIEKA	373	
Db	426	SEBESSSEBETKQWVAITTKP-KATAKAALSLFPAKQAPQSRDSSSDSSSEBEEBK	484	
QY	374	ANPSGKKLKIPOPVIEAPGASKICIGPCCHVAOPDSVYCNDCILKHAHAATMKRFLSSGK	433	
Db	485	TSKSAVKK-----KPKQVAGGAAP-----SKPASA-----KKGKAESSNSSSD	523	
QY	434	EQKPAPKEKMKMK-PEKPSLPKCGAQAQIGIKISSVHKGPAPKEKETTIVKAVVVPARSBAL	492	
Db	524	DSSEBEEKLKGKSPRQAPKANGTSALTQNGKAAKNSEBEEBCKKAAVWVSKSGSL	583	
QY	493	GK-----PAACESSTP	503	
Db	584	KCKKONEAAKEAETP	598	

RESULT 14

TRDN_HUMAN	STANDARD;	PRT;	728 AA.
ID	TRDN_HUMAN		
AC	Q13061;		
CD	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Triadin.		
GN	Triadin.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Skeletal muscle;		
RX	MEDLINE=96061957; PubMed=7588753;		
RA	Tagaki N.L., Evrie H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,		


```
FT CONFLICT 75 76 QA -> PT (IN REF. 1).
FT CONFLICT 100 100 S -> R (IN REF. 2).
FT CONFLICT 106 106 A -> S (IN REF. 1).
FT CONFLICT 115 115 L -> V (IN REF. 1).
FT CONFLICT 621 621 K -> R (IN REF. 1).
FT CONFLICT 633 633 MISSING (IN REF. 1).
FT CONFLICT 642 642 K -> R (IN REF. 1).
FT CONFLICT 654 654 T -> S (IN REF. 1).
FT CONFLICT 664 664 N -> S (IN REF. 1).
FT CONFLICT 670 670 P -> A (IN REF. 1).
FT CONFLICT 698 698 N -> S (IN REF. 1).
FT CONFLICT 712 719 AYLEPTVT -> BALIMPTS (IN REF. 1).
FT CONFLICT 727 727 A -> V (IN REF. 1).
FT CONFLICT 752 752 R -> K (IN REF. 1).
FT CONFLICT 782 782 S -> N (IN REF. 1).
FT CONFLICT 817 817 A -> T (IN REF. 1).
FT CONFLICT 823 823 S -> G (IN REF. 1).
FT CONFLICT 833 833 N -> I (IN REF. 1).
FT CONFLICT 837 841 SGKSA -> NMGT (IN REF. 1).
FT CONFLICT 849 849 A -> T (IN REF. 1).
FT CONFLICT 857 857 D -> E (IN REF. 1).
FT CONFLICT 862 862 MISSING (IN REF. 1).
FT CONFLICT 888 888 I -> V (IN REF. 1).
FT CONFLICT 949 950 SK -> NR (IN REF. 1).
FT CONFLICT 954 955 SA -> NT (IN REF. 1).
FT CONFLICT 959 960 LT -> PA (IN REF. 1).
FT CONFLICT 964 964 T -> A (IN REF. 1).
FT CONFLICT 968 968 T -> I (IN REF. 1).
FT CONFLICT 972 974 ISA -> ASAST (IN REF. 1).
FT CONFLICT 979 979 T -> A (IN REF. 1).
FT CONFLICT 1038 1040 TGA -> SGS (IN REF. 1).
FT CONFLICT 1045 1045 T -> A (IN REF. 1).
FT CONFLICT 1068 1068 MISSING (IN REF. 1).
SQ SEQUENCE 1101 AA; 121231 MW; A703F6C98C16B471 CRC64;
```

```
Query Match 5.6%; Score 167; DB 1; Length 1101;
Best Local Similarity 38.3%; Pred. No. 0.0637;
Matches 31; Conservative 16; Mismatches 20; Indels 14; Gaps 4;

QY 269 LYCICQPHN-NRFMICCDRCCEWFGDCVGESEAGRLLENGEDYICPNCITILQVQDE 327
DB 6 VYCVCLPYDVTFRFMEICDCKDWFHSCVGVVEEBAPDI----DIYHCPNC-----EK 55

QY 328 THSETADQOEAKWR---PGDA 345
DB 56 THGKSTLKKRTWHRKGPQA 76
```

Search completed: April 30, 2004, 08:54:27
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 17:22:09 ; Search time 5297 Seconds
(without alignments)
4598.600 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGDPSNEEAPKAIKPTS.....RSFWIAIPWACPGIGVRAALC 562

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool_p/US09787016/runat 30042004 083702 4968/app_query.fasta_1.711
-DB=GenEmbl -PMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016@cgn 1.1 5265 @runat 30042004 083702 4968 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Result No.	Score	Query Match	Length	DB ID	Description
1	2989	100.0	2610	6	BD244712 Genes enc
2	2989	100.0	2610	6	AX023369 Sequence
3	2989	100.0	2725	9	BC014489 Homo sapi
4	2989	100.0	2767	9	BC004237 Homo sapi
5	2981	99.7	2772	6	AX878224 Sequence
6	2981	99.7	2772	6	BD157094 Primer fo
7	2981	99.7	2772	9	AK002127 Homo sapi
8	2811	94.0	2407	9	BC000770 Homo sapi
9	2112.5	70.7	2602	10	MU238332 Mus muscu
10	2112.5	70.7	2867	6	BD244713 Genes enc
11	2112.5	70.7	2867	6	AX023370 Sequence
12	2081.5	69.6	4906	10	AK129117 Mus muscu
13	1575	52.7	6692	9	AB002331 Human mRN
14	1372.5	45.9	272404	2	AC108337 Rattus no
15	1342	44.9	104663	10	AL732560 Mouse DNA
16	1281	43.2	2428	5	BC060442 Xenopus 1
17	1282	42.9	160241	9	HS88547 Human DNA
18	1101	36.8	28519	9	HSU563514 Human DNA
19	918.5	30.7	221547	2	AC119716 Rattus no
20	817.5	27.4	69252	2	AC101519 Mus muscu
21	549	18.4	681	6	AX869935 Sequence
22	549	18.4	681	6	BD149997 Primer fo
23	518	17.3	69252	2	AC101519 Mus muscu
24	465.5	15.6	198133	2	EX296530 Danio rer
25	419	14.0	2161	3	AK114522 Clona int
26	267.5	8.9	61204	2	AC017132 Drosophil
27	267.5	8.9	161601	3	AC007594 Drosophil
28	267.5	8.9	225655	3	AB003695 Drosophil
29	252	8.4	112387	2	AC098841 Magnapor
30	237.5	7.9	2867	9	HS8807333 Homo sapi
31	232	7.8	5975	6	AX329603 Sequence
32	232	7.8	6256	6	AR338778 Sequence
33	232	7.8	6936	9	D87685 Human mRNA
34	232	7.8	6948	6	AX210673 Sequence
35	232	7.8	6948	9	AF091622 Homo sapi
36	232	7.8	7142	9	HS8808416 Homo sapi
37	231.5	7.7	2452	9	AF149758 Homo sapi
38	231.5	7.7	2487	9	AB031069 Homo sapi
39	228.5	7.6	2718	5	BC056775 Danio rer
40	214.5	7.2	2308	9	BC015733 Homo sapi
41	214.5	7.2	2320	6	AX086854 Sequence
42	214.5	7.2	2320	9	HS8801830 Homo sapi
43	214.5	7.2	2331	9	BC014940 Homo sapi
44	214.5	7.2	68050	8	EX842636 Neurospor
45	211	7.1	3298	8	AK111180 Oryza sat

ALIGNMENTS

RESULT 1

BD244712 2610 bp DNA linear PAT 17-JUL-2003
LOCUS Genes encoding for the human and murine death
DEFINITION inducer-obliterator-1.
BD244712
ACCESSION BD244712.1 GI:33054482
VERSION
KEYWORDS JP 2002526040-A/1.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 2610)
AUTHORS Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, E. and Martinez, P.
TITLE Genes encoding for the human and murine death inducer-obliterator-1
JOURNAL Patent: JP 2002526040-A 1 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Homo sapiens (human)
PN JP 2002526040-A/1
PF 20-AUG-2002
PR 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5, 17-SEP-1998 US 60/100873 P1
CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC A61P35/00, A61K38/00, A61K45/00, A61K48/00, A61P3/10, A61P17/00, PC
A61P29/00,
PC A61P35/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12Q1/02, C12Q1/68, G01N33/15, G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00, A61K37/02
CC This gene is referred to in the application as Human Death CC
Inducer
CC Obliterator Gene 1.
CC It has now been named by the Human Gene
Nomenclature Committee
CC as
CC Death-Associated Transcription Factor (DATF-1) PH Key
Location/Qualifiers
FT source 1..2610
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..2610
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 1,228-118 Length: 2610
Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-787-016A-3 (1-562) x BD244712 (1-2610)
QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
DB 265 ATGGACGCAAAAGGCGACCGAGCAATGAGGAGGCACTTAAGGCCATCAACCCACCAAGC 324
QY 21 LysGluPheArgHisThrTroGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
DB 325 AARGAGTTTCAGGAAACATGGGGTTTTCAGAGGACCACTATGCGCCAGCGAGAGGGCCCA 384
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 385 GGGGACCGGGAGGCTGACCCACTGGAGCGCGCCACCCACACAGCAGCAGCTGGGCTGTGCC 444
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 445 CTGGCGCGCAGTGGGAGGCGAGCCCAAGCGCACTGAGCGCGTGGAGCACTTCTCTGACCAAT 504

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 505 GCGCGGCGCGCGCGAGGAGCATGCTGTCTCCCTGGAGGATCTGTGGAGCCACG 564
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 565 TCTTGCCTCCCGCCACAGAGCGCCGAGACAGCTCTCCGAGGGCAGCGTGGAAAGCGCTTCTGAG 624
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 625 ACCAGAGCGCGCCCGCCAGTCTGTCTCCACAGCTGTGTGAAGGAACGACGAGCTCTTCTGAA 684
QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 685 AAGTGAAGAGAGGGGATGACCAAGATGACACTCCGATAGTACAGCGATGGCTTCTGACC 744
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
DB 745 TTGAAGAGCTTCAGATCGCTCTCCGAGGAGCGGACAGGAGGCCACTGAGAGGCC 804
QY 181 LeuLysGlyLeuGlnSerArgLeuArgLysArgArgGluGlnGlyProAlaGluThr 200
DB 805 CTGAAGAGGATCCAGAGTCCGCTCCGGAAGAGCGCGGAGGAGGGTCCGCGCAGACT 864
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
DB 865 GTGGGCTCCGAGCGCAGTGACACTGTGGAGGGGCTCTGCGCCAGTAAGCAGGAGCCGAG 924
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
DB 925 AACGATCAGGGGGTGTGCTCCAGGCTGGGAAGATGACAGAGAGATTAAGTGGAGGGA 984
QY 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
DB 985 AAGCGGCTCAGGACATCAAGATGAGGCGCTGGAGACTTGGCGCCACCGAGCCGTGA 1044
QY 261 CysGluGlyThrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
DB 1045 TGTGAGGGTTACGACCCCAACGCTCTGTATTGCTATTCGCCAGCTCACAACACAGG 1104
QY 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSer 300
DB 1105 TTTATGATTTGCTGTGACCGCTGTGAGATGGTTTCATCGCGATTCGTGGGCAATTCT 1164
QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
DB 1165 GAGGCTCGAGGAGGCTTTTGGAAAGGATGGGGAAGACTATATCTGCCCAAACTGCACCC 1224
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
DB 1225 ATTCTGCAAGTGCAGGATGAGATCTCATTCAGAAACCGCAGATCAGCAGGAGCTAATGG 1284
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
DB 1285 AGACCTGAGATGCTGATGCGACCGATTGTACAAGTATAGGACATATATCTGCCCAAACTGCACCC 1344
QY 361 SerGluAspGlnGlyIleLysGlyArgGluLeuAlaAlaAsnProSerGlyLysLys 380
DB 1345 AGCGAAGCAAGGGATAAAGGGTAGAATTTGAGAAAGCTGCAATCCAAATCCAGTGGCAGAAG 1404
QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
DB 1405 AAACCTCAAGATCTTCCAGCTGTGATAGAGCGCGCTGGTGGCTCAAAATGTATTGGCCCC 1464
QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerHisAspCysIleLeuLys 420
DB 1465 GGGTGTCTGTCGTCGGCGGAGCCGACTCGGTGTACTGCACTAATGATGATCTCTCTCAAA 1524
QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
DB 1525 CAGCCCGCAGCGACATGAGATTTCTTAAGCTCAGTAAAGACAGAGCCAAAGCCCTAAA 1584

QY 441 GluLysMetIleMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 DB 1585 GAAAGATGAAGATGAAGCCAGAGAGCCAGCTCTCCGAAATGCGGTCTCAGGCAGGT 1644
 QY 461 IleIleValSerValHisLysArgProAlaProGluLysGluThrThrValLys 480
 DB 1645 ATTAAATCTCTCTGTGCAAGAGACAGCTCCAGAAAAAAGAGACCAAGGTGAAG 1704
 QY 481 LysAlaValValProAlaArgSerGluAlaLeuGluLysGluAlaAlaCysGluSer 500
 DB 1705 AAGGCAGTGTCTCTCTGCGAGTGAAGCACTCGGGAAGCAAGCAGCTTGTGAGAGC 1764
 QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 DB 1765 AGCACCCGCTGCGGCGAGCGATCAATTAATGCAATGCAAGCAAGCAAGCAAGCTGCT 1824
 QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGluValGluLysLeuLeuAsp 540
 DB 1825 GCTCTCTGCGCTCACTGTGTATTAATGTATGATACCTAGGCTTGGCTCTCTGGAC 1884
 QY 541 ProSerArgSerPheThrIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 DB 1885 CCCTCCCGTCTCTTCTGATAGCATCCCTGCGGCTGTCCAGGCTGGGAGTTCAGCT 1944
 QY 561 LeuCys 562
 DB 1945 TTGTGT 1950
 RESULT 2
 LOCUS AX023369 2610 bp DNA linear PAT 15-SEP-2000
 DEFINITION Sequence 1 from Patent WO015787.
 ACCESSION AX023369
 VERSION AX023369.1 GI:10183781
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Leonardo, B.; Martinez, P.; Alonso, C.M.; Domingo, D.G. and Grandien, A.
 TITLE Genes encoding for the human and murine death in
 ducer-oblierator-1
 JOURNAL Patent: WO 0015787-A 1 23-MAR-2000;
 LEONARDO ESTHER (ES); MARTINEZ PEDRO (ES); CONSEJO SUPERIOR
 INVESTIGACION (ES); DOMINGO DAVID GARCIA (ES); MARTINEZ ALONSO
 CARLOS (ES); BANNERMAN DAVID GARDNER (GB); GRANDIEN ALF (SE)
 FEATURES
 source
 1..2610
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="This gene is referred to in the application as
 Human Death Inducer Obliterator Gene 1. It has now been
 named by the Human Gene Nomenclature Committee as
 Death-Associated Transcription Factor (DATF-1)."
 ORIGIN
 Alignment Scores:
 Pred. No.: 1..22e-118 Length: 2610
 Score: 2989.00 Matches: 562
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-787-016A-3 (1-562) x AX023369 (1-2610)
 QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
 DB 265 ATGGACGACAAAGCGGAGCCGAGCATGAGAGGCGACCTAAGGCGCATCAAAACCCAGGC 324
 QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGlyAla 40

DB 325 AAAGAGTTGAGGAAACATGGGGTTTTGAGAGGACCACTATCGCCAGGAGAGGGCGCA 384
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 DB 385 GGGAGCGGGAGGCTGACCCACTGCGAGCGGCCACCCACACAGCAGCTGGGCTGTCTC 444
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 445 CTCGCGCGCAGTGGAGGCGCCCAAGCGCACTGAGCGCGTGGAGCGATTCTCTGACCAT 504
 QY 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 505 GCGCGCGCGCGCGGAGGAGGAGCATGCTGTCTCTCTGAGGATTCGTGTGAGCCCGC 564
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 565 TCTGCTCCCGCCACAGACCCCGAGCAGCTCCGAGGCGCGTGGAAAGCGCTTCTGAG 624
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
 DB 625 ACAGAGAGCGCGCCCGCTGCTCTCCAGCTGTGAGGAGACGACCGCTCTCTGAA 684
 QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 685 AAGGTGAAAGGAGGGGATGACCCAGATGACCTCCGATAGTGACAGCGATGCGCTGACC 744
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 DB 745 TTGAAGAGCTTCAGATCGCTTCCAGAGAGCGGAAACAGAGCCCTCTGAGAGGCC 804
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200
 DB 805 CTGAAGGAGTATCCAGATGCTCTCGGAGAGAGCGCGGAGGAGGTCCCGCCGAGACT 864
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 DB 865 GTGGGCTCCGAGCCAGTGACACTGTGGAGGGCGTCTCTCCAGTAAAGCAGGAGCCGAG 924
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 DB 925 AAGCATCAGGGGTGTGTCTCCAGCTGGGAAGATGACAGAGAGATTAAGTTGGAGGGA 984
 QY 241 LysAlaAlaGluAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 DB 985 AAGCGGCTCAGGACATCAAGATGAGGAGCCCTGGAGACTTGGGCGCAGCCGAGCCTGAA 1044
 QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 DB 1045 TGTGAGGGTTACGACCCCAACCCCTGTATTGCAATTTGCGCGCAGCTCACACACAGG 1104
 QY 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
 DB 1105 TTTATGATTGTCTGTCACCGCTGTGAAGATGGTTTCATGCGCATTTGTCTGGGCAATTTCT 1164
 QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 DB 1165 GAGGCTCGAGGGAGGCTTTTGGAAAGGATGGGAGAGACTATATCTGCCCAACTGCACC 1224
 QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLysTrp 340
 DB 1225 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGCTAAATGG 1284
 QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 DB 1285 AGACCTGGAGATGCTGATGGCCCGCATTTGTACAAGTATAGGAACAATAGAGCAGAGTCT 1344
 QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 DB 1345 AGCGAGACCAAGGAGTAAGGGTAGAATTGAGAAAGCTGCAATTCGAAGTGGCAAGAAG 1404
 QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400

Db 1405 AAACCTCAAGATCTCCAGCCCTGTGATAGAGCGCCTGTGCTCCAAATGATTGGCCCC 1464

Qy 401 GlyCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLys 420

Db 1465 GGGTGGCTGTCACTGGCGGCGAGCCGACTCGGTGCTACTGCGAGTATGACTGTATCTCTCAA 1524

Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440

Db 1525 CACGCGCGAGCGCAATGAAGTTCTTAAGCTCAAGTAAAGACACAGAGCCAAAGCTTAA 1584

Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460

Db 1585 GAAAGATGAAGATGAAGCAGAGAGAGCCAGTCTTCGGAATGCGGTGCTCAGGAGGT 1644

Qy 461 IleLysIleSerSerValHisLysAspProAlaProGluLysLysGluThrThrValLys 480

Db 1645 ATTAATATCTCTCTGACACAGAGACAGCAGCTCCAGAAAAGAGAGACACAGTGAAG 1704

Qy 481 LysAlaValValValProAlaArgSerGluAlaGluGlyLysGluAlaAlaCysGluSer 500

Db 1705 AAGCAGTGTGTGCTCTGGCGGAGTGAAGCAGTCTGGGAGGAGAACGCTTGTGAGAGC 1764

Qy 501 SerThrProSerThrAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAla 520

Db 1765 AGCAGCGCGTGTGGCGGAGGAGTCAATACATGAGTAAAGCCAGGAAAGACTGCT 1824

Qy 521 AlaProSerProSerLeuLysLysCysMetTyHisLysGlyValGlyLeuLeuLys 540

Db 1825 GCTCCCTCGCGTACTGTGTATAATGATATATATATATATATATATATATATATATAT 1884

Qy 541 ProSerArgSerPheThrIleAlaIleProThrAlaCysProGlyLeuGlyValAlaAla 560

Db 1885 CCTCCCGTCTTCTTGTGATGAGCATCCCTCGGCTGTCCAGGACTGGAGTTCAGCT 1944

Qy 561 LeuCys 562

Db 1945 TTGTGT 1950

RESULT 3

BC014489

LOCUS

DEFINITION Homo sapiens death associated transcription factor 1, transcript variant 2, mRNA (cdna clone MGC:23216 IMAGE:4901057), complete cds.

ACCESSION BC014489

VERSION BC014489.1 GI:15680266

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2725)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullighy, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Heintz, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smalley, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

12477932
2 (bases 1 to 2725)
Strausberg, R.
Direct Submission
Submitted (17-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 34 Row: n Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.

FEATURES
source

Location/Qualifiers
1..2725
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:23216 IMAGE:4901057"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH MGC 17"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
1..2725
/genes="DAF1"
/notes="synonyms: DIO1, FLJ11265, KIAA0333"
/db_xref="LocusID:11083"
/db_xref="MIM:604140"
238..1926
/codon_start=1
/product="death associated transcription factor 1, isoform
a"

gene

CDS

/protein_id="AAH14489.1"
/db_xref="GI:15680267"
/db_xref="LocusID:11083"
/translations="MDDKDPGNEAPKAIKTSKEFKTWGFRRTTIARREGAGDAE
ADPLEPPQQQLGLSRSGQPKHTEVQFLIARERGRSMPSVLEDSGEPTSC
PATDAASEGSVESASSTRSQSASTAVKRPASSEKVGKGDHDDTSDSDGLT
LKEQLNLRKREQRTPEPLKQISRLAKRREBPATVSGSEADTVEGVLPSKQE
FENDQGVSOAGKDRSLKGAQIDKDEBPGDLGRPKPCBEGDPPNLYCICQEP
HNNRFICCDCEBWFHPCVGISEARGELRNGEDYICPNCTILQVQDETHSETAP
QOAKWRPGDAGTCTCTGTTGKSESDQIGKRIKAANPSGKKKLIPOFVTRAP
GASKICPGCHVAPQDSVVCNDICILKHAATMKFLSSGKEQKPKPKMKMKPKP
SLPKCAAGIKISSVHKPAPKPKTKTKVAVVVPARSEALQKKAACSSPTSPWASD
HYNVAKPKTRAPSPSLLYKCMWHLGVLLDPSRSFWIAIPWACRGLGVALC"
1048..1197
/notes="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"

misc_feature

1048..1197
/notes="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FYVE domains"
/db_xref="CDD:pfam00628"

ORIGIN

Alignment Scores:

Pred. No.: 1.28e-118 Length: 2725

Score: 2989.00 Matches: 562
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-787-016a-3 (1-562) x BC014489 (1-2725)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLeuLysProThrSer 20
Db 238 ATGACGACAAAGCGCGACCGCGCAATGAGGAGCCACTTAAGGCCATCAAAACCCACGCG 297
QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 298 AAGAGTTTCAAGAAACATGGGTTTTCGAGGACCACTATCGCCAGCGAGAGGGCGCA 357
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
Db 358 GGGGACCGGAGGCTGACCCACTGAGCGCCGACCCCGCCACGACGACGAGCTGGGCCCTGTC 417
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 418 CTGGCGCGCAGTGGGAGGAGCCCGCCAGCGCACTGAGGCGCTGAGCAGTTCTTGACGCTT 477
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 478 GCGCGCGCGCGCGAGGAGCATGCTGCTCTCTCTGAGGATTTCTGGTGGAGCCACG 537
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 538 TCTTGGCCCCCGCACAGACCGCGACAGCTTCCGAGGCGAGCTGAGAAAGCGCTTCTGAG 597
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 598 ACCAGAGCGCGCCCGCCAGTCTGCTTCCACAGCTGTGGAAGAACGACGAGCTCTTCTGAA 657
QY 141 LysValLysGlyGlyArgAspAsnHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 658 AAGGTGAAGAGGAGGAGATACCCAGTGCACCTCCGATAGTACAGCGATGGGCTTGACC 717
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Db 718 TTGAAGAGCTTCAGATCGCTTTCGAGGAGCGGGAACGAGGAGCCACTGAGAGGCGCC 777
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGlnGluProAlaGluThr 200
Db 778 CTGAAGAGGATCCAGAGTCCGCTCGGAAGAACGCGCGGAGGAGGCTCCCGCGGAGACT 837
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 838 GTGGGCTCCGAGGCGAGTGACACCTGTGAGGAGGCTCTGCGCCAGTAGAGGAGGCGCGAG 897
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Db 898 AACGATCAGGGGGTGTGTGTCAGGCTGGGAAGATGACAGAGAGATGATGTTGGAGGGA 957
QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db 958 AAGCGCGCTCAGGACATCAAGATGAGGAGCTTGAGACTTGGGCGCCAGCGAAGCCCTGAA 1017
QY 261 CysGluGlyTyrosProAsnAlaLeuTyrosLysCysArgGlnProHisAsnAsnArg 280
Db 1018 TGTGAGGTTTACGACCCCGCCCTGTATTGCTATTTGCGCCGAGCTTCAACACAGG 1077
QY 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSer 300
Db 1078 TTTATGATTTCTGTGACCGCTGTGAAGATGGTTTATGCGGATGTGGGGCATTTCT 1137
QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspFyrIleCysProAsnCysThr 320
Db 1138 GAGGCTCGAGGAGGCTTTTGAAGGATGCGGAGAGCTATATCTGCCCAACATGCAAC 1197
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340

Db 1198 ATTCTCAGTGCAGGATGAGACTCATTCCAGAAACGCGAGATCAGCAGGAGCTAATGG 1257
QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlySer 360
Db 1258 ACACCTGGAGATGCTGATGGCAGCCGATTTGACAGTATAGGACCAATAGCAGCAAGTCT 1317
QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
Db 1318 AGCGAGACCAAGGNTAAGGTTAGAAATGAGAAAGCTGCAATCCAAATGGTGGCAAG 1377
QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
Db 1378 AAACCTCAGATCTTCCAGCTGTGATAGAGGCGCTGCTGCTCAAAATGTATTGGCCCC 1437
QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrosSerAsnAspCysIleLeuLys 420
Db 1438 GGTGCTGTCTGTCGCGGCGGCGGCTCGGTACTGAGTAATGACTGTATCTCTCAA 1497
QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 440
Db 1498 CACGCGCGCAGCAATGAATTTCTTAGCTCAGGTAAAGAACAGAGGCAAGCCCTAA 1557
QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
Db 1558 GAAAGATGAGATGAAGCGAGAGCCGAGTCTTCCGAAATGCGGTGCTCAGGCGGT 1617
QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
Db 1618 ATTAATAATCTCTTCTGTGCAAGACAGCAGCTCCAGAAAAAAGAGACCAAGTGAAG 1677
QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
Db 1678 AAGCGAGTGGTGTCTCTCGCGAGTGAAGCACTCGGAGGAAGCAGCTTGTGAGAGC 1737
QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrosAlaValLysProGluLysThrAla 520
Db 1738 AGCAAGCGCTCTGCGGCGAGCGATCAATTAATGAGTAAAGCAGAAAAAGACTGCT 1797
QY 521 AlaProSerProSerLeuLeuTyrosCysMetTyrosHisLeuGlyValGlyLeuLeuAsp 540
Db 1798 GCTCCCTCCGCGCTCCTGTGTATTAATGTATGATCACTAGGGGTGGCCCTCCGAGC 1857
QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
Db 1858 CCTCCCGTCTTCTGTGATAGCCATCCCTGCGGCTTCCAGGACTGGGAGTTGAGCT 1917
QY 561 LeuCys 562
Db 1918 TTGTGT 1923

RESULT 4

BC004237

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC004237 2767 bp mRNA linear PRI 12-NOV-2003
Homo sapiens death associated transcription factor 1, transcript
variant 1, mRNA (cDNA clone MGC:896 IMAGE:2959493), complete cds.

BC004237

BC004237.2 GI:33872756

MGC.

Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2767)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klauser, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zensberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hong, P.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalls, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 2767)
 Strausberg, R.
 Direct Submission
 Submitted (01-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:13278980.
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 2 Row: 0 Column: 16
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 18375620.
 Location/Qualifiers

1..2767
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:896 IMAGE:2959493"
 /tissue type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC 15"
 /lab_host="DH10B-R"
 /note="Vector: pORF7"
 1..2767
 /gene="DATF1"
 /note="synonyms: DIO1, FLJ11265, KIAA0333"
 /db_xref="LocusID:11083"
 /db_xref="MIM:604140"
 282..1970
 /codon_start=1
 /product="Death associated transcription factor 1, isoform
 a"

/protein_id="AAH04237.1"
 /db_xref="GI:13278981"
 /db_xref="LocusID:11083"
 /translation="MDDKDPNBEAPKALPTSKFRFTTAKRGAGDAE
 ADLPPEPQQQLSLRSGQPKETVEVDFLTABRGERSMPVSLSDSPTSC
 PADATASGSEVSASSTRSPQASNAVKERPASSKYGGDDHDTSDSDSLT
 LKELQNRKKKEQBPETPLKGIQSRLLKRRKREGPARTVGEASDSTVEGVLPKQE
 PNDQGVVQAGDKREGKLEGAQDINDPEPDLGRPKPECEGYDPAIYICRQP
 HNNRPMICDRCEWFGHDCVGISRARGLLERNGEYICFNCTTILQVDETHSETAD
 QQAKWRPGDCHVPTSTIGTTEOKSSSDOGIKGRIKKAANPSGKKKIKIIPQVTEAP
 GASKICIGPCDCAVQSDSVYCSNDICILKHAATMKFLSSGKEQKPKPKMKMKPEK
 SLPKGAQAGIKLSVHKPEPAKPKETTVKAAVVVPAPSEALGKAEACESTPSWAD
 HYNVAKPEKTPAAPSFLYKCMYHLGVGLLDPSFNLIPWACFGVGVVALLC"
 1092..1241
 misc_feature
 /note="PHD; Region: PHD-finger. PHD folds into an

interleaved type of Zn-finger chelating 2 Zn ions in a
 similar manner to that of the RING and FYVE domains"
 /db_xref="CDD:pfam00628"

ORIGIN

Alignment Scores:

Pred. No.: 1..3e-118 Length: 2767
 Score: 2989.00 Matches: 562
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x BC004237 (1-2767)

QY 1 MetAspAspLysGlyAspProSerAsnGluLalaProLysAlaLysProThrSer 20
 DB 282 ATGACCAACAAGCGGACCCGAGCAATGAGGAGGACCTAAGGCCATCAACACCCACGAGC 341
 QY 21 LysGluPheArgLysThrTpGlyPheArgThrThrileAlaLysArgGluGlyVala 40
 DB 342 AAAGAGTTTCAGAAACATGGGGTTTTCGAAGGACCACTATCGCCAAAGCAGAGGCGCA 401
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnGlnLeuGlyLeuSer 60
 DB 402 GGGACGCGGAGGCTGACCCACTCGAGCGCGCACCCCCACAGCAGCAGCTGGGCTTGTCC 461
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrile 80
 DB 462 CTGGGCGCAGTGGAGGCGGCCCAAGCGCACTGAGCGCGTGGAGCAGTCTCTGACCAT 521
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 522 CGCGGCGCGCGCGGAGGAGGAGCATGCTGTCTCTCTCTGGAGGATCTCTGTGAGCCACG 581
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 582 TCCTGCCCCCGCACAGACCGCGAGCAGCTTCCGAGGCGCGCTGGAAAGGCGCTTCTGAG 641
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 DB 642 ACCAGAGCGCGCGCCCACTGCTTCCACAGCTGTGAAGGAACACGAGCCCTCTTCTGAA 701
 QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 702 AAGGTGAAGAGGAGGAGATGACCAACATGACACCTCCGATAGTACAGCGGATGGCTGACC 761
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 DB 762 TTGAAGAGCTTCAGATCGCTTCCAGGAAGCGGGAACAGGAGGCCCACTGAGAGGCC 821
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgArgGluGluGlyProAlaGluThr 200
 DB 822 CTGAAGGCGATCCAGAGTCGCTTCCGAAAGAGCGCGGAGGAGGGTCCCGCGAGACT 881
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 DB 892 GTGGGCTCCAGAGCGAGTGAACATCTGTGGAGGCGCTCTCTCCCACTAGCAGAGGCCCG 941
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 DB 942 AACGATCAGGGGGTGTGTGTCACGCTGGGAAAGATGATCAGAGAGAGTAGTAAGTTGAGG 1001
 QY 241 LysAlaAlaGlnAsnLysAspGluLysProGluGlyAspLeuGlyArgProLysProGlu 260
 DB 1002 AAGCGCGCTCAGGACATCAAGATGAGGAGCGCTGGAGCTTGGCGCGGAGCGGAGCTGAA 1061
 QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnArg 280
 DB 1062 TGTGAGGGTACGACCCCAACGCGCTGTATTGCTTTTCCGCCAGCCTCACAACAACAGG 1121
 QY 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspValGlyIleSer 300

QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 901 GTGGGCTCCAGAGGCGAGTGACACCTGTGAGGCGCTCTCTGCCAGTAAAGCAGGAGCCGAG 960

QY 221 AsnAspGlnGlyValValSerGlnAlaGlyValSerAspArgGluSerLysLeuGlnGly 240
 Db 961 AACCATCAGGGGGTGTGTCCAGCTGGGAAGATGACAGAGAGTAAGTTTGAGGGA 1020

QY 241 LysAlaAlaGlnAspIleLysAspGlnGluProGlyAspLeuGlyArgProLysProGlu 260
 Db 1021 AAGCGGCTCAGGACATCAAGATGAGGAGCTCAAGACTTTGGCCGACGACGAGCTGAA 1080

QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 Db 1081 TGTGAGGGTTACGACCCCAAGCGCTGTATTGCAATTTGCCGCCAGCCCTCAACAACAGG 1140

QY 281 PheXetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300
 Db 1141 TTTATGATTTCTGTGACCGCTGTGAGAATGGTTTCATGGCGATTTGTGGGCATTCT 1200

QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyIleAspTyrIleCysProAsnCysThr 320
 Db 1201 GAGGCTCAGGAGGCTTTTGGAAAGGATGGGAAGACTATATCTGCCCAACATGACAC 1260

QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaIleTyr 340
 Db 1261 ATTTGCAAGTGCAGGATGAGACTCATTCAGAACCGCGATTCAGCAGGAGCTAATATGG 1320

QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 Db 1321 AGACCTGAGATGCTGTATGGCACCGATTGTACATATAGGAACAATAGAGCAGAACTCT 1380

QY 361 SerGluAspGlnGlyIleGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 Db 1381 AGCGAAGACCAAGGATAAAGGTAGAAATTGAGAAGCTGCAAATCCAAAGTGCAGAG 1440

QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1441 AAACTCAAGATCTTCCAGCGTGTATAGAGGCGCTGTGCTCAAAATGTATTGGCC 1500

QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLys 420
 Db 1501 GGGTCTGTCTGCTGCGCGAGCCGACCTGGGTACTGCGATGATGATGTATCTCAAA 1560

QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1561 CACGCGCAGCGACATGAGTTTCTAAGCTCAGGTAAAGAACACAGAGCCAAAGCTTAA 1620

QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1621 GAAAAGATGAATGAAGCAGAGAGCCAGCTCTTCCGAAATGCGGTCTCAGGAGGT 1680

QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1681 ATTAAATCTTCTGTGCAAGAGACAGCTCCAGAAATAAAGAGACACAGTGAAG 1740

QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlnSer 500
 Db 1741 AAGGCAGTGTGTCTCTGCGGAGTGAAGCACTCGGAGAGAGAGAGCTTGTGAGAGC 1800

QY 501 SerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 Db 1801 AGCAGCGCTGTGCGGCGAGCGATCACATTTCAATGATCAATGAAAGAGAGAGAGCTGCT 1860

QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 Db 1861 GCTCCCTGCGCTCACTGTTGTATTAATATGATATATACCTAGGAGGTTGGCTCTCGAC 1920

QY 541 ProSerArgSerPheTyrIleAlaIleProTyrAlaCysProGlyLeuGlyValAlaAla 560
 Db 1921 CCTCCCGCTTCTTCTGGATAGCCATCCCTCTGGGCTGTCTCAGGAGTGGAGTTCAGCT 1980

QY 561 LeuCys 562
 Db 1981 TTGTGT 1986

RESULT 6
 BD157094
 LOCUS
 DEFINITION
 Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION
 BD157094
 VERSION
 BD157094.1 GI:27862852
 JP 2002191363-A/11937.
 KEYWORDS
 Homo sapiens (human)
 SOURCE
 Homo sapiens
 ORGANISM
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002191363-A 11937 09-JUL-2002;
 HELIX RESEARCH INSTITUTE
 COMMENT
 OS Homo sapiens (human)
 PN JP 2002191363-A/11937
 PD 09-JUL-2002
 PF 28-JUL-2000 JP 2000280990
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
 SAITO,
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
 KEIICHI NAGAI, TETSUJI OTSUKI
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10,
 C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key
 Location/Qualifiers
 (301)..(1986).
 FT CDS Location/Qualifiers
 i. 2772
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores: 2.84e-118 Length: 2772
 Pred. No.: 2981.00 Matches: 561
 Score: 99.82% Conservative: 0
 Percent Similarity: 99.82% Mismatches: 1
 Best Local Similarity: 99.73% Indels: 0
 Query Match: 6 Gaps: 0
 DB: 6

US-09-787-016A-3 (1-562) x BD157094 (1-2772)

QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
 Db 301 ATGACGACAAAGCGACCCGAGCAATGAGGAGGACCTAAGGCCATCAACCCACCCAGC 360

QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 Db 361 AAACAGATTTCAGAAAAAATGCGGGTTTTCGAAAGGACCACTATCGCCAGCGAGGAGGCGCA 420

QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 421 GGGGACGCGGAGGCTGACCCACTGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTGCC 480

QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 481 CTGCGGACGCAATGGGAGGAGCGCCAGCGCACTGAGCGCGTGGAGGAGCTTCTGACCAT 540

QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 541 GCGGCGCGCGCGCAGGAGGAGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600

PATDAETASBGSVESASBTRSGPOSASTAVKPERPASSEKVKYGGDDHDDTSDSDGUT
 LKLEONLRREBQBTERRPLKIQSRLRKKRREBPAETVGSSEASDTEBGLVPSKOE
 PENQGVVSOAGKDDRESKLEAKAODIKDEPDLGRPKPECEGYDPNLYCICROP
 HNNRPMICDCEEFHEDGCVGISEARGLLERNGEDYICNPCTILOVDDETHSETAP
 QOENRREDADGDTCTSIGTHIKKSSDQGIKRIIEKANPSCKKLKLPQVLEAP
 GASKCTPGCHVAPQDSVYICNSDCILKRAATVFLSSGKEQPKPKERKKMKRDEK
 SLPKGQAGGKIISVHRPAPAKETITVKAIVVPKSEALGACCESSFPWSASD
 HNNVAVPEKTAAPSPSLLYKQVHLGVLDPSPSEFWIALPMACPGGLGVAALC"

ORIGIN

Alignment Scores:

Pred. No.: 2,846-118 Length: 2772
 Score: 2981.00 Matches: 561
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.73% Indels: 0
 Gaps: 9

US-09-787-016a-3 (1-562) x AK002127 (1-2772)

Qy 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 301 ATGGACGACAAAGGCGACCCGAGCAATGAGGAGGACCTTAAGGCCATCAACCCACCCAGC 360
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
 Db 361 AAAGAGTTCAGGAACAATGGGGTTTCGAGAGCACCTATGCCACAGCGAGAGGGCGCA 420
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 421 GGGGACGGGAGGCTGACCACTGGAGCGCGCACCCACACAGCAGCAGCTGGGGCTGTCC 480
 Qy 61 LeuArgArgSerLysArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 481 CTGGCGCGCAGTGGGAGGAGCGCCAAAGCGCACTGAGCGGTGGAGCAGTTCTGACCAFT 540
 Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 541 GCGGCGCGCGCGGAGGAGGAGCATGCTGCTCTCCCTCGAGGATTTCTGGTGAGCCACG 600
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 601 TCTGTCGCCCGCACAGACCCGAGACAGCTTCGAGGGCAGCGTGGAAGCGCTTCTGAG 660
 Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 661 ACCAGAGCGGCCCCCGCTCTGCTTCCAGCTGTGAGAGACGACCCCTCTTCTGAA 720
 Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 721 AAGGTCAAGAGGAGGAGGATGACCACTGACATCCCTCGATAGTGACAGCGATGGCTGACC 780
 Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGlnProThrGluArgPro 180
 Db 781 TTGAAAGAGCTTCAGAAATCGCTTCGAGAGCGGGAAACAGGAGCCCACTGGAGGCGCC 840
 Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200
 Db 841 CTGAAAGGATCCAGAGTCCCTGCGGAAGAGCGCGGAGGAGGAGGCTCCCGCCGAGACT 900
 Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 901 GTGGGCTCCGAGGCCCATGACATGTGGAGGGGCTCTGCCCATGTAGCAGGAGGCCGAG 960
 Qy 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 Db 961 AACGATCAGGGGCTGTGTGTCAGGCTGGGAAAGATGACAGAGAGAGTAGTAAAGTTGAGGCA 1020
 Qy 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 Db 1021 AAGCGCGCTCAGACATCAAGATGAGGAGCCCTGAAGACTTGGGCCGAGCGAGCCCTGAA 1080
 Qy 261 CysGluGlyTyraAspProAsnAlaLeuTyraCysIleCysArgGlnProHisAsnAsnArg 280

Db 1081 TGTGAGGGTTACGACCCCAACGCCCTGTATGCAATTTGCCGCCAGCCTCAACACACAGG 1140
 Qy 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSer 300
 Db 1141 TTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTCATGCGAFTGTGTGGCATTTCT 1200
 Qy 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCysThr 320
 Db 1201 GAGGCTCGAGGGAGCGCTTTTGGAAAGGANTGGGGAAGACTATATCTGCCCAACTGCACC 1260
 Qy 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlnAlaLysTrp 340
 Db 1261 ATTCTGCAAGTGCAGATGAGACTCATTCAGAAACCGCAGATCAGCAGGAAGCTAAATGG 1320
 Qy 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 Db 1321 AGACCTGGAGATGCTGATGGCACCGATTGTACAAATATAGAAACAATAGACAGAAGTCT 1380
 Qy 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 Db 1381 AGCGAAGACCAAGGANTAAAGGTAGAAATTGAGAAAGCTGCANATCCAAATGGCAAGAG 1440
 Qy 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 Db 1441 AAACTCAAGATCTTCCAGCCTGTGATAGAGCGCCTGCTGCTCAAAATGTATTGGCCCC 1500
 Qy 401 GlyCysCysHisValAlaGlnProAspSerValTyraCysSerAsnAspCysIleLeuLys 420
 Db 1501 GGGTGTCTGTACGTGGCGCGACGCCGACTCGGTGTACTCCAGTATGACTGTATCTCTCAA 1560
 Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1561 CACGCCGCGAGCACAAATGAAGTTTCTAAGCTCAGTAAAGAACAGAAAGCCAAAGCCTAAA 1620
 Qy 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyValaGlnAlaGly 460
 Db 1621 GAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCGAAATTCGGTGCTTCAGCGAGT 1680
 Qy 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1681 ATTAATCTCTCTCTGTGCACAGAGACCACTCCAGAAAAAAGAGACACACAGTGAAG 1740
 Qy 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGluSer 500
 Db 1741 AAGCGAGTGGTGGTCCCTGCGCGAGTGAAGCATCTCGGGAAGGAGAGCTGTGAGAGC 1800
 Qy 501 SerThrProSerTrpAlaSerAspHisAsnTyraAsnAlaValLysProGluLysThrAla 520
 Db 1801 AGCAGCGCTCTGTGGCGCGGATCACATTTACATGTCAGTAAAGCCAGAAAGACTGCT 1860
 Qy 521 AlaProSerProSerLeuLeuTyrllysCysMetTyrlHisLeuGlyValGlyLeuLeuAsp 540
 Db 1861 GCTCCCTCGCGTCACTGTGTATAAATGTATGTATATCACTAGGGGTGGGCTCTCTGGAC 1920
 Qy 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 Db 1921 CCTCCCGCTTCTTCTGTGATAGCATCCCTGGGCTGTCCAGGACTGGGAGTTGCAGCT 1980
 Qy 561 LeuCys 562
 Db 1981 TTGTGT 1986

RESULT 8

BC000770

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC000770 2407 bp mRNA linear PRI 12-NOV-2003
 Homo sapiens death associated transcription factor 1, transcript
 variant 3, mRNA (cdna clone MGC:3257 IMAGE:3506207), complete cds.
 BC000770
 BC000770.2 GI:33990873
 MGC
 Homo sapiens (human)
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 2407)	
REFERENCE	
AUTHORS	
Strausberg,R.D., Collins,F.S., Wagner,L.H., Derge,J.G., Klausner,R.D., Felling,S.D., Shemen,C.M., Schuler,G.D., Altchul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.P., Bhat,N.K., Hopkins,R.F., Jordan,H.K., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,F., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,P.D., Mullaby,S.J., Bosak,S.A., McGowan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,J., Helton,B., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grinchmod,J.J., Smetz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.B., Gennerich,A., Schein,J.E., Jones,S.J. and Marra,M.A.	
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
JOURNAL	
MEDLINE	
22388257	
12477932	
2 (bases 1 to 2407)	
Strausberg,R.	
Direct Submission	
Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
NIH-MGC Project URL: http://mgc.nci.nih.gov	
On Aug 20, 2003 this sequence version replaced gi:12653952.	
Contact: MGC help desk	
Email: cgabbs@mail.nih.gov	
Tissue procurement: DCTD/BTP	
cDNA Library Preparation: Rubin Laboratory	
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)	
DNA Sequencing by: Institute for Systems Biology	
http://www.systemsbioology.org	
contact: amadansystemsbioology.org	
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting	
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov	
Series: IRAL Plate: 7 Row: 1 Column: 2	
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.	
Location/Qualifiers	
1..2407	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="MGC:3257 IMAGE:3506207"	
/tissue_type="kidney, renal cell adenocarcinoma"	
/clone_lib="NIH MGC_14"	
/lab_host="DH10B-r"	
/note="Vector: pOT57"	
1..2407	
/gene="DATF1"	
/note="synonyms: DIO1, FLJ11265, KIAA0333"	
/db_xref="LocusID:11083"	
/db_xref="MIM:604140"	
201..1835	
/codon_start=1	
/product="death associated transcription factor 1, isoform b"	
/protein_id="AAH00770.1"	
/db_xref="GI:12653953"	
CDS	
FEATURES	
SOURCE	
gene	
CDS	

```

/db_xref="locusID:11083"
/translation="MDDKGDPSNEAPKAIKPTSKFRKTWGFRTTTAKREGADAE
ADPLEPPPOQLGLSLERSGQPKRTVERFOFLIARRGRMRSPVLSDEGEPTPS
PATDAETASGESVSASTRSGPSATVKRPAESPASKEVGSDDDHDDTSDSDGLT
LKEIQLNRLLKREQEPTEPRLQGIQSLRLLKRRREBPAPATVGSSEADTVGVGLPSKQE
PENNQVVOAGKDDREKLKGAADIKDEEPGLDGRPKPCEGVDPNALYICICROP
HNNEVVICDCEEPHGDYVIGSEARGLLERNGEDYICNPCTILOQVDETHSETAD
QORAKWRPGDADGTDCTSIGTIEQKSSSEDQIKRIEKAAMPSSKKKLKLPQVLEAP
GASKICGQCHVAPQVPSVYCSNDILKRLKAAATMFLSGKGQKSPKPKRMKPKKPKP
SLPCKGAGTKISGVHREPAPEKBTITVKAAVVPAPARSEALGKEAAECESSTPSSWASD
ENTYNAPEKTAAPSPSLLYKCSGKLYLSLPSLIA"
misc_feature
1011..1160
/note="PHD; Region: PHD-finger. PHD folds into an
interleaved type of Zn-finger chelating 2 Zn ions in a
similar manner to that of the RING and FIVE domains"
/db_xref="CDD:pfam00628"

ORIGIN

Alignment Scores:
Pred. No.: 2407
Score: 2811.00 Matches: 536
Percent Similarity: 36.43% Conservative: 5
Best Local Similarity: 95.54% Mismatches: 12
Query Match: 94.04% Indels: 8
DBs: 9 Gaps: 2

US-09-787-016A-3 (1-562) x BC000770 (1-2407)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaAlaIlysProThrSer 20
DB 201 ATGGACGACGAAGCGCGACCCGAGCAATGAGGAGGACCATTAAGCCATCAACACCCACG 260
QY 21 LysGluPheArgLysThrTIGlyPheArgArgThrThrIleAlaIysArgGluGlyAla 40
DB 261 AATGAGTTTCAGGAACACATGGGTTTTTCGAGGACCACTATCGCCACGAGGCGCGCA 320
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 321 GGGGACGCGGAGGCTGAGCCACTGAGGAGCGGCCACCCACACAGCAGCAGCTGGGCGCTTGC 380
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 381 CTGCGCGCGAGTGGGAGGAGCGCCCAAGCGCATCTGAGCGGTGGAGCGAGTTCTTCACCAT 440
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 441 GCGCGCGCGCGCGGCGAGGAGGACATGCTGTCTCCCTGGAGGATTCGTGTGAGCCACG 500
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 501 TCTCTCCCGCCACAGACCGCGAGACGCTCCGAGGCGCAGCGTGGAAAGCGCTTCTGAG 560
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 561 ACCAGAGCGCGCCCCAGTCTGTCTCCACAGCTGTGAAGGACGACCAAGCTCTTCTGAA 620
QY 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 621 AAGGTGAAGAGGAGGAGATGACCAACATGACACCTCCGATAGTAGCAGCGATGCGCTGACC 680
QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
DB 681 TTGAAAGAGCTTCAGAAATCGCTTCGAGGAGCGGGAACAGGACCCCACTGAGAGGCC 740
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGluGlyProAlaGluThr 200
DB 741 CTGAAGAGGATCCAGAGTGCCTGCGAGAGAGCGCGCGGAGAGGGTCCCGCCGAGACT 800
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
DB 801 GTGGGCTCCGAGGCCCHGTGACACTGTGAGGGCGCTCTCCGCCAGTACAGAGGAGCCGAG 860
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240

```

LOCUS MWU238332 2602 bp mRNA linear ROD 09-JUL-1999
Mus musculus mRNA for death inducer-oblierator-1 (Dio-1).

DEFINITION
ACCESSION AJ238332
VERSION AJ238332.1 GI:5457402
KEYWORDS death inducer-oblierator-1; Dio-1 gene.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 Garcia-Domingo, D., Leonardo, E., Grandien, A., Martinez, P., Albar, J.P., Ipiñua-Belmonte, J.C. and Martinez-A.C. Dio-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development
Proc. Natl. Acad. Sci. U.S.A. 96 (14), 7992-7997 (1999)
99324176
MEDLINE 1039335
PUBMED 2 (bases 1 to 2602)
AUTHORS Garcia-Domingo, D.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Garcia-Domingo D., Department of Immunology and Oncology, Centro Nacional de Biotecnología/CSIC, Campus Cantoblanco UAM, Ctra. de Colmenar Viejo Km 16, Madrid, 28049, SPAIN

FEATURES
source Location/Qualifiers
1. .2602
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cell_line="WOL-1"
/cell_type="pre-B1"
1. .2602
/gene="Dio-1"
232. .>2073
/function="apoptosis inducer"
/evidence="experimental"
/product="death inducer-oblierator-1"
/protein_id="CA848401.1"
/db_xref="GI:5457403"
/db_xref="GOA:Q8C9B9"
/db_xref="GDB:Q9WV00"
/translation="MDKGLHNEAPKAIKPTSKFRKWTGFRRTTAKREGAGDTEADSEQQPQHNLSLRSGPQKPRTERVEFLTTVRRGKNVPSVLESDSEPTSTVTDVETASGEVSSEIRSGPVSLSLKEHPASSEKAGGEEEDTSDSDGLTLKE LQNLRRKREBPVERSLSQNLKRGKREDEPREAGKPECEVIDPDNALYICICRPHNN SQPVSQSTDDIENQLEGKATQNTENPREAGKPECEVIDPDNALYICICRPHNN RFMCCDCREWFHGDVGISEARGELLENEDYICNPCTILQVQDENGSAFNEDQ SGCRVAGDGTCTGTGTGKSGEDQIGRIEKAANPSGKKGLKIPVVEAPGAP KCTGPGCSSVAQPDVSVNDCILKHAATWRLSSGKQKPKPKPKPKPKPKPK KCSVQVGIKISSVHKRLASERENPVKWLASRSETSGKEAECSTPSWSDHNTN AVXPKPKKTLALSPILLKCTTHPKAGFPQPSHLGSCIGLSRVLGLVLIVASS SLPSRSKYQASGPGVFLPSLWSLSGFLASCVGLMLEAISYFSFRP#"

ORIGIN
Alignment Scores: 1.94e-81 Length: 2602
Pred. No.: 2112.50 Matches: 422
Score: 81.40% Conservative: 33
Percent Similarity: 75.49% Mismatches: 94
Best Local Similarity: 70.68% Indels: 11
Query Match: 10 Gaps: 7
DB: 10

US-09-787-016A-3 (1-562) x MWU238332 (1-2602)

QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
DB 232 ATGGATGATGATAAGGGACCTGAGCAATGAGGAGCCACCCAGGCTATCAACCCACCACT 291
QY 21 LysGluPheArgLysThrTyrPheArgThrThrIleAlaLysArgGluGlyAla 40

861 AACGATCAGGCGGTGTGTCCTCCAGCTGGGAAAGATCAGAGAGATAGTTGGAGGA 920
241 LysAlaAlaGlnAspIleLysAspGluProGlyAspLeuGlyArgProLysProGlu 260
921 AAGCGCGTCAGGACATCAAGATGAGGAGCTGGAGACTTGGGCCACCGAAGCTGAA 980
261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAspArg 280
981 TGTGAGGCTTACGACCCACAGCCCTGTATGTCATTTGGCGCCAGCTCACAACACAGG 1040
281 PheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300
1041 TTTATGATTGCTGTGACCGCTGTGAAGATGGTTTCATGCGGATTTGTGGGCAATTTCT 1100
301 GluAlaArgGlyArgLeuLeuLeuArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
1101 GAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAGACTATATCTGCCCAAACTGCACC 1160
321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyr 340
1161 ATTCTGACGTCAGGATGAGACTATTCAGAAACGCGCAGATCAGCAGGAGCTAAATGG 1220
341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
1221 AGACCTGAGATGCTGTATGTCACCGCATTTGTACAAGTATAGCAACATAGACGAGTCT 1280
361 SerGluAspGlnGlyIleLysGlyValGlyIleGluLysAlaAlaAsnProSerGlyLysLys 380
1281 AGCGAAGCCAGGATTAAGGGTAGAATTGAGAGCTGCAATCCAGTGGCAGAGAG 1340
381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
1341 AAACCTCAAGATCTTCCAGCTGTGTAGAGAGCGCTGCTGCTCAAAATGTTATGGCCCC 1400
401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLysLys 420
1401 GGGTCTGTCTGCTGCGCGCAGCCGACCTCGGTGTAATCGAGTAAGTACTGTATCTCAAA 1460
421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
1461 CACGCGCGCAGCGACCAATGAAGTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCTAAA 1520
441 GluLysMetLysMetLysProGluLysProSerIleuProLysCysGlyAlaGlnAlaGly 460
1521 GAAAAGATCAAGATCAAGCCAGAGAGCCAGCTCTTCGAAATGCGGTGCTCAGGACGT 1580
461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
1581 ATTAAATCTCTCTGTGCACAGAGACAGCTCCAGAAAAGAGAGACACACAGTGAAG 1640
481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
1641 AAGGCAGTGTGCTGCTGCTGCGCGAGTGAAGCACTCGCGAAGAGAGAGCTTGTGAGAGC 1700
501 SerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
1701 AGCAGCGCTGTGCGGCGAGCCGATTCACATTAATGCAATGCAATGCAATGCAATGCAAT 1760
521 AlaProSerProSerIleuLysLysCys-----MetTyrHisIleuGlyVal 536
1761 GCTCCCTCGCGCTCACTGTGTATATAATCTCTGTTAGTATTTGTTATTTCTTCTTCC 1820
537 GlyLeuLeuAsp-----ProSerArgSerPheThrIleAlaIleProThrAla 552
1821 AGTCTGATTGCTAGCACCACCTGCGCGCGCGCGCCACATCCACCTCTGCTGACATGAT 1880
553 Cys 553
1881 TGT 1883

Db 292 AAGGAGTTCAGGAAACCTGGGGTTTTTCGAGAAACACGAGTTCCTCCAAACGTGAGGTGCA 351
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
Db 352 GGAGACAGGAGCGGAGCCAGTGTAGAGCAACCA-----CAGCAGCATACCTCTCC 405
Qy 61 LeuArgSerGlyArgGlnProProLysArgThrGluArgValGlnGlnPheLeuThrIle 80
Db 406 CTGGCCGCGAGTGGACGCGAACCAACGACTGTGAGAGGTAGAGAGTTCCTTACACG 465
Qy 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 466 GTTGGCGCGGAGGAGAAAGAAAGTGTCCGCTGTCTGGAGGATTCAGTGTAGCGCCACA 525
Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 526 TCTTCCACAGTCACTGATGTGGAGACAGCTTCCGAGGGAGCGTGTGAAGACAGTCTTGAG 585
Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 586 ATCAGAGTGGCCCTGTACTGTACTCTTAGGG---AAAGAACATCTCTCTCTGAA 642
Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 643 AAGCCAAAGAGGTGAAGAGGAGGAGAGACACCTCTGACAGTGCAGTGTATGCGCTTACG 702
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgGluArgGluGlnGlnProThrGluArgPro 180
Db 703 TTGAGAGACTTTCAGAAACCCCTTCGAGAAAGCGAGAGAGAACCTGTGGAGAGTCC 762
Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluGluProAlaGluThr 200
Db 763 CTGAGAGCAGTCAAGATCGCTCAGAGAGAGAGAGAGAGAGAGAGAGATTCGCGCAACT 822
Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
Db 823 ---GGAGTTCCAATAGGAGTCCGAGCAGAGACAGACCTCTCTGTAGAGAGAGCCT 879
Qy 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
Db 880 GAGCTAGTCAGGAGCCAGTGTCCAGTCAGAGACAGACAGTACATAGATAATCAGTTGGAA 939
Qy 240 GlyValAlaGlnAlaGlnPheLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
Db 940 GGGAGGCGACTCAGGGAATATACAGAGAAACCCAGGAGAGCGGCGCAACCAAGCCT 999
Qy 260 GluCysGluGlyTyrosProAsnAlaLeuTyrosLysCysArgGlnProHisAsnAsn 279
Db 1000 GAGTGTGAGTTTACGACCCCAATGCTGTACTGTATCTGCGCCAGCTCACAACAC 1059
Qy 280 ArgPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIle 299
Db 1060 AGGTTTATGATCTGCTGTGATCGGTGTAGAGGTGTCTCCATGTGTGTGGGTATT 1119
Qy 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrosLysCysProAsnCys 319
Db 1120 TCTGAGCCCGAGCGCGGCTCTCGAAAGAGACGGGAGAGACTACATCTGCCCAATTCG 1179
Qy 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
Db 1180 ACCATTTCAGTGCAGTGTAGACAAACCGTAGCCACCAATGAGCAGGACTCTGGG 1239
Qy 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlnGlnLys 359
Db 1240 TGCAGATCTGTGGTGTGTGATGGCAGACTGTGCACAGCATAGGACGTAGAGCAGAG 1299
Qy 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
Db 1300 TCCGGAGACACAGGCGCATAAAGGGTAGGATTGAGAGGCGAGCAACCCAGCGGCAAG 1359
Qy 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
Db 1360 AAAAAGACTCAAGATATTCCAGCCCTGTGTAGAGGCTCTCTGTCTCTTAATGATTCGC 1419

Qy 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrosSerAsnAspCysIleLeu 419
Db 1420 CTGGGTGTTCAGTGTAGCAGAGCTGACTCTGTGTATGTAGTAATGACTGCTATTC 1479
Qy 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
Db 1480 AAACAGCAGCAGCTTACCATGAGATTTCTAAGTTCAGGTAAAGAACAAACCAACCC 1539
Qy 440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
Db 1540 AAGGAAAGGTTCAGACGAGGAGGAGAAAGTTCAGTCTTCCAAATPCCAGTGTTCAGGTG 1599
Qy 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVal 479
Db 1600 GGGATTAAATCTCTCTGTGTGCAGAGAGACTAGCGTCAGAGAAAGGAGAAACCCAGTG 1659
Qy 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysGlu 499
Db 1660 AAGAAA---GTGATGTGCTTCCAGGAGTGTAGACTTTCGGAAGGAGGAGCGCTGTAG 1716
Qy 500 SerSerThrProSerThrAlaSerAspHisAsnTyrosAlaValLysProGluLysThr 519
Db 1717 AGCAGCACCATCTCTGGGCAAGTGTACCACTACATGCTGTAAAGCAGAGAGCCCA 1776
Qy 520 AlaAlaPro-----SerProSerLeuTyrosCysMetTyrosHisLeuGlyVal 536
Db 1777 GAGAGCCCACTGCACTCTGCGCCACCTTATGAGTAAATGTACGTATCACCCAAAGCT 1836
Qy 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaLeuProTrpAlaCysProGly 555
Db 1837 GGTCTCCAGGCGCCCTCCCATC---ATCTGGGTGGCT---GCTGGGGCTGTCTAGGA 1888

RESULT 10
BD244713 2867 bp DNA linear PAT 17-JUL-2003
LOCUS Genes encoding for the human and murine death inducer-oblierator-1.
DEFINITION
BD244713
VERSION BD244713.1 GI:33054483
KEYWORDS JP 2002526040-A/2.
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Alonso, C.M., Domingo, D.G., Grandien, A., Leonardo, E. and Martinez, P.
TITLE Genes encoding for the human and murine death inducer-oblierator-1
JOURNAL Patent: JP 2002526040-A 2 20-AUG-2002;
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
COMMENT OS Mus sp. (murine)
PN JP 2002526040-A/2
PD 20-AUG-2002
PF 10-SEP-1999 JP 2000570314
PR 10-SEP-1998 SE 9803069-5, 17-SEP-1998 US 60/100873 PI
CARLOS MARTINEZ ALONSO, DAVID GARCIA DOMINGO, ALF GRANDIEN, PI
ESTHER LEONARDO,
PI PEDRO MARTINEZ
PC C12N15/09, A61K38/00, A61K45/00, A61K48/00, A61P3/10, A61P17/00, PC
A61P29/00,
PC A61P35/00, A61P37/06, A61P43/00, C07K14/47, C07K16/18, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12Q1/02, C12Q1/68, G01N33/15, G01N33/50//C12P21/08, PC
C12N15/00,
PC C12N5/00, A61K37/02
CC Genes encoding for the human and murine death inducer- oblierator-1
FH Key 1.2867 Location/Qualifiers
FT source /organism= 'Mus sp. (murine)'.
FT Location/Qualifiers
1.2867 /organism= 'Mus sp.'

/mol_type="genomic DNA"
/db_xref="taxon:10095"

ORIGIN

Alignment Scores:

Pred. No.: 2,14e-81 Length: 2867
Score: 2112.50 Matches: 422
Percent Similarity: 81.40% Conservative: 33
Best Local Similarity: 75.49% Mismatches: 94
Query Match: 70.68% Indels: 11
DB: 6 Gaps: 7

US-09-787-016A-3 (1-562) x BD244713 (1-2867)

```
Oy 1 MetAspAlaGluAlaAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 232 ATGGATGATAAAGGCGACCTCGACCAATGAGGAGCACCACCAAGGCTATCAACCCACCACT 291
Oy 21 LysGluPheArgLysThrTidGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
Db 292 AAGGAGTTTCAGGAAACCTGGGGTTCGAGAACCAACGATGCGAAGCTGAGGGTGCA 351
Oy 41 GlyAspAlaGluAlaAspProLysGluProProProGlnGlnGlnLysGlyLeuSer 60
Db 352 GGAGACACGAGGCGGACCCACAGTCAGCAGCAACCA-----CAGCAGCATAACCTCTCC 405
Oy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 406 CTGCCCGCAGTGGACGGCACCAACACTACTGAGAGGTGAGAGTTCTTACACAG 465
Oy 81 AlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 466 GTTCGGCGCCGAGGAGAAAGAAATGCGCGTGTCCCTGGAGATTCCAGTCAGGCCACA 525
Oy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 526 TCTTCCACAGTCACATGATGAGACAGCTTCCGAGGGGAGGCTTGAAGCAGTCTCTGAG 585
Oy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 586 ATCAGAGTGGCCCTCTATCTGCTCTCTAGGG---AAAGAACATCTGCTCTCTGAA 642
Oy 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 643 AAGGCAAAAGGAGGTGAGAGGAGAGACACCTCTGACAGTGACAGTGAATGCGCTTACG 702
Oy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Db 703 TTGAAGAACTTCAGAACCGCTTCGAGAAAGCGCAGACAGACCACTCTGAGAGGTCC 762
Oy 181 LeuLysGlyIleLysSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200
Db 763 CTGAGAGGCAGTCAGAAATCGCTGAGAAAGAGCGCAGAGAGAAATCTGCGCGAAATCT 822
Oy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
Db 823 ---GGAGTGTCCAAATAGCAGTCGCGAGCAGGACAGACCTCTCTGTAAGCAGAGCTCT 879
Oy 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
Db 880 GAGGCTAGTCAGGACCACTGCTCCAGTCAGACAGACAGATGACATAGAAATCAGTTGGAA 939
Oy 240 GlyLysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
Db 940 GGGAGGCGCAGTCAGGGAATCAGAGGAAACCCCGAGGAAGCGGCGCAACCAAGCTCT 999
Oy 260 GluCysGluGlyTyraAspProAsnAlaLeuTyraIleCysArgGlnProHisAsnAsn 279
Db 1000 GAGTGTGAGTTTACGACCCCAATGCCCTGTACTGCTATCTGCGCGCAGCTCACAACCAAC 1059
Oy 280 ArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIle 299
Db 1060 AGGTTTATGATCTGCTGTCATCGGTGAGGAGTGGTTCCATGGTGACTGTGTGGGTATT 1119
```

```
Oy 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCys 319
Db 1120 TCTGAGCCCGAGCGCGCTCTCGAAAGAGAAAGGAGAGACTACTACTCTCCCAATATGC 1179
Oy 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
Db 1180 ACCATTTCGAGTGCAGGATGAGCAACACGAGTGGCCACCAATGAGCAGGACTCTGGG 1239
Oy 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
Db 1240 TGCAGATCTCTGGGTGCTGATGGCAGACACTGCACAGCATGAGGACAGTAGAGCAGAG 1299
Oy 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
Db 1300 TCCGAGAGAGACCGGCGCATTAAGGCTAGGATGAGAGGCGCAGCAACCCAGCGGCAAG 1359
Oy 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
Db 1360 AAAAACTCAGATATTCAGCTCTCTAGAGGCTCTCTGCTCTCTCTCTCTCTCTCTCTCT 1419
Oy 400 ProGlyCysCysHisValAlaGlnProAspSerValTyraCysSerAsnAspCysIleLeu 419
Db 1420 CCTGGGTGTTCCAGTGTAGCAGCCCTGACTCTCTGTATGTCAGTATGACTGCTATCTC 1479
Oy 420 LysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
Db 1480 AAACACCGCAGCTTACCATGAGATTTCTAAGTTTCAGTAAAGAACCAAAACCAAAACC 1539
Oy 440 LysGluLysMetLysMetLysProGlyLysProSerLeuProLysCysGlyAlaGluAla 459
Db 1540 AAGGAAAGGTCAAGACGAAAGGCTTCAAGTTCTTCAAAATGTCAGTGTTCAGGTG 1599
Oy 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVal 479
Db 1600 GGGATTAATCTCTCTGTGTCACAGAGACTAGCGTCAGAGAAAGGAAAGCCAGCTG 1659
Oy 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
Db 1660 AAGAAA---GTGATGCTGGCTTCAGGAGTGCAGCTTCTGGAAAGGAGGAGCTGTGAG 1716
Oy 500 SerSerThrProSerTrpAlaSerAspHisAsnTyraAlaValLysProGluLysThr 519
Db 1717 AGCAGCACACCTCTCTGGCAGTGCAGCAACTACATGCTGTGAGCAGCAGAGAGCA 1776
Oy 520 AlaAlaPro-----SerProSerLeuLeuTyrlsCysMetTyrlsLeuGlyVal 536
Db 1777 GAGAAGCCCACTGCACTCTCGCCCACTTATGAGTAAATGATGATATCACCCAAAGGCT 1836
Oy 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
Db 1837 GGTTCACAGGCCCTCTCCATC--ATCTGGGTGGCT---GCTGGGGCTGTCTAGTA 1888
```

RESULT 11

AX023370 2867 bp DNA linear PAT 15-SEP-2000
LOCUS Sequence 2 from Patent WO0015787.
DEFINITION AX023370
ACCESSION AX023370.1 GI:10183782
VERSION
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Leonardo, B., Martinez, P., Alonso, C.M., Domingo, D.G. and Grandien, A.
TITLE Genes encoding for the human and murine death in
JOURNAL ducer-obliterators-1
Patent: WO 0015787-A 2 23-MAR-2000;
LEONARDO ESTHER (ES); MARTINEZ PEDRO (ES); CONSEJO SUPERIOR
INVESTIGACION (ES); DOMINGO DAVID GARCIA (ES); MARTINEZ ALONSO
CARLOS (ES); BANNERMAN DAVID GARDNER (GB); GRANDIEN ALF (SE)
Location/Qualifiers

```

source      1. .2867
            /organism="Mus sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:10095"

ORIGIN

Alignment Scores:
Pred. No.:      2,14e-81      Length:      2867
Score:          2112.50      Matches:      422
Percent Similarity: 81.40%      Conservative: 33
Best Local Similarity: 75.49%      Mismatches: 94
Query Match:    70.68%      Indels: 11
DB:             Gaps: 7

US-09-787-016A-3 (1-562) x AX023370 (1-2867)

QY 1 MetAspAspIysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
DB 232 ATGCATGATAAAGGACACCTGAGCAATGAGAGCAACCAAGGCTATCAACCCACCACT 291
QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrLeuAlaLysArgGluGlyAla 40
DB 292 AAGGAGTTTCAGGAAACCTCGGGTTTTCGAGAAACCAACGATTCGCAAGGTGAGGTCGA 351
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnLeuGlyLeuSer 60
DB 352 GAGACACGAGGCGGACCCCTGAGCAGCAACCA-----CAGCAGCATTAACCTCTCC 405
QY 61 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 406 CTCGCGCGCAGTGGAGCGGCAACCAACCAACGACTGAGAGGGTGAAGAGTTTCTTACACG 465
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 466 GTTCGCGCGCGAGGAAAGAGATGTCCGCTGTCTCCGAGGATTCAGTGAAGCCCA 525
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 526 TCTTCCACAGTCTCATGTGAGAGCAGCTTCCGAGGCGGCGTTGAAAGCAGCTCTGAG 585
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 586 ATCAGAGTGGCCCTGTATCTGACTCTTAGGG---AAAGAACATCTGCTCTTCTGAA 642
QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 643 AAGGCAAAAGAGGTGAAGAGGAGGAAGACACCTCTGACAGTGAACAGTATGCGCTTACG 702
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgGluGlnGluProThrGluArgPro 180
DB 703 TTGAGAGACTTCAGAACCCCTTCGAGAAAGCAGAGCAGAGACTCTGAGAGAGTCC 762
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluProAlaGluThr 200
DB 763 CTCGAGGCGAGTCAAGATCGCTGAGGAAGAGCGCAGAGAGAGATTCTGCGGAAACT 822
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
DB 823 ---GGAGTGTCCAAATAGGAGTGGCGGACAGCAGACCTCTCTGTAGCAGGAGCCT 879
QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
DB 880 GAGCGTAGTCAGGACCCAGTGTCCAGTCAGACAGACAGATGCATAGAAAATCAGTTGGA 939
QY 240 GlyValAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
DB 940 GGGAGCGGACTCAGGGAATATACAGGAAACCCCGAGGAGCGGCAACCAAGAGCCT 999
QY 260 GluCysGluGlyTyrosProAsnAlaLeuTyrosLysCysIleCysArgGlnProHisAsnAsn 279
DB 1000 GAGTGTGAGTTTACGACCCCAATGCGCTGTATGTATCTGCTGCGCCGAGCTTCAACAAC 1059
QY 280 ArgPheMetIleCysCysAspArgCysGluGluThrTrpPheHisGlyAspCysValGlyIle 299

```

RESULT 12

AK129117

LOCUS

AK129117

DEFINITION

Mus musculus mRNA for mKIRA0333 protein.

ACCESSION

AK129117.1

VERSION

GI:37359897

KEYWORDS

FLI CDNA.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

1

AUTHORS

Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S.,

Sage, Y., Nagase, T., Ohara, O. and Koga, H.

Prediction of the coding sequences of mouse homologues of KIAA

gene: III. the complete nucleotide sequences of 500 mouse

KIAA-homologous cDNAs identified by screening of terminal sequences

of cDNA clones randomly sampled from size-fractionated libraries

AK129117 4906 bp mRNA linear ROD 21-NOV-2003

Mus musculus mRNA for mKIRA0333 protein.

AK129117.1 GI:37359897

FLI CDNA.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S.,

Sage, Y., Nagase, T., Ohara, O. and Koga, H.

Prediction of the coding sequences of mouse homologues of KIAA

gene: III. the complete nucleotide sequences of 500 mouse

KIAA-homologous cDNAs identified by screening of terminal sequences

of cDNA clones randomly sampled from size-fractionated libraries

JOURNAL DNA Res. 10 (4), 167-180 (2003)
 MDLINE 22977043
 PUBMED 14621295
 REFERENCE 2 (bases 1 to 4906)
 AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-2003) Hisashi Koga, Kazusa DNA Research
 Institute, Laboratory for Genome Informatics; 2-6-7
 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
 COMMENT The CREATE program supported by Japan science and technology
 corporation; cDNA full insert sequencing; Kazusa DNA Research
 Institute; cDNA library construction, clone selection and 5'- &
 3'-end one pass sequencing.

FEATURES
 source
 1..4906
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="mbgl6015"
 /tissue_type="brain"
 /dev_stage="adult"
 /note="vector:modified pBC SK+"

gene
 1..4906
 /gene="mKIAA0333"

CDS
 <224..3829
 /gene="mKIAA0333"
 /note="CDS is predicted by in silico analysis. Start codon
 is not identified."
 /codon_start=1
 /evidence="not experimental"
 /product="mKIAA0333 protein"
 /protein_id="BAC97927.1"
 /db_xref="GI:37359898"
 /translation="PVVFPAPKFOAFSGFTGMDKGLSNEEAPKAIKPTSKPRFKT
 WGPRTTAKREGAGTEVDSEQQPHNLISLRSGQPKRTERVEFLATVRRCK
 KNPVLSDESSEPTSTVDVETASEGVSSEIRSGPVSLSGKHPASSERAKGG
 EEEEDTSDSGDLTKSLQNLRRKQEPVRSLSGQNLRRKREEDSAAETGSV
 QIGSAEQPLCKQPEASQGVSDTDIENQLEKATQNTENPREAGKPKPEC
 EYVDNALYICRQPHNRPMICCDRCBEMFHGDCVGISEARGLLENRGHDYICPN
 TILQVDTGTSATDEQSGCRSCVADGTDCTSIGTVOKGSGEQGIKGRLEKAAPS
 GKXKLIQPVVEAPGAPKICGPGCSSVAQPDVSYCNDILKHAATWRLSSGKEQ
 KTKPKKVTXPKPKSLPKCSVOVGIKIISVHVHKLASKEPNPVKVMLASRSTSGK
 EACSSSTSWASDHNINAVKPEKPKPTALSPILLSSKMDRDRVETWAAVTIPK
 KALPSALVIGQTSRPNVVKLPYPYNNAGAKPAIKLPGSGFKGTIPRRVPSATLS
 GTSARAGPTWNAASKLPGSAVGVTRKPMFSAVPAAPGLRGPVSPAPSQN
 SQIRQNIIRSLKEILKLVNDSDDLIMTENEVGIKALHIEKEMFLFQVTDNRYSKY
 RSIIMNLKDPKNOGLFHRVREEISLAKLVRMKPEELVSKLSMWTEKPTKSVIESRT
 KLLNSKNTKPTETIPMEDSPVSDSEKQESVRAAPEKSAAPLLDVFSSMLKDTT
 SOHRAHLFDLNCIKITGQVPSSEDEPAKCKLKSASKEDEPKPHDSPPNVAQNTA
 DEGIADTLFENASEPDPSTESINQKRCFPESPGDSHPPESSLGGLSPSASGSGV
 VTTVTMGDRDPTALSGSCTVTASMAHLNDSQASSETKLDIMKPAITSAVVPKSLAK
 PSSSPDPRLSYPPSPSISERSPEEGDTTLPLSLNTIWKGFIMQSVAKFVTKAYP
 VSGCLDLSLEDLPDTHIGGRAPKTVWDVYVGLKLSVSKLCIRFHPATSEEVAY
 ISLSYFSSRGRGFGVANNRHHVLDYLIPLSAKDPVPSKLIIPFGPKGHPVGR"

ORIGIN

Alignment Scores:
 Pred. No.: 7,72e-80 Length: 4906
 Score: 2081.50 Matches: 412
 Percent Similarity: 82.80% Conservative: 31
 Best Local Similarity: 77.01% Mismatches: 83
 Query Match: 69.64% Indels: 9
 DB: 10 Gaps: 6

US-09-787-016a-3 (1-562) x AK129117 (1-4906)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 278 ATGGATGATAAGGGCACCTGAGCAATGAGGAGCACCCCAAGCTATCAACCCACCACT 337
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40

Db 338 AAGGAGTTCAGGAAAACTCGGGGTTTTTCGAAGAACACCGATTGTCACAAACCTGAGGGTGCA 397
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
 Db 398 GGGAGACACGGAGGTGGACCCCGAGTGAGCAACCA-----CAGCAGCATAACTCTTCC 451
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 452 CTGGCCCGCAGTGGAACGCAACCAACCTACTGAGAGGGTAGAAGAGTTCTTACCACG 511
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 512 GTTCGGCGCCGAGGAGAAAGATGTCCCGTGTCCCTGGAGGATTCCAGTGAGCCACACA 571
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 572 TCTTCCACAGCTCACTGATGTGGACAGCTTCCGAGGGGAGCGTTGAAAGCAGTTCTGAG 631
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 632 ATCAGAAAGTGGCCCTGTATCTGACTCTCTTAGGG---AAAGAACATCTCGCTCTTCTGAA 688
 QY 141 LysValLysGlyArgAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 689 AAGCAAAAGAGGTGAAGAGGAAAGAGACACCTCTCAGCTGACAGTATGGCTTACG 748
 QY 161 LeuLysGluLeuGlnAanArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
 Db 749 TTGAAGGAACCTTCAGAACCGCTTCGAGAAAGCGAGAGCAAGAACCTGTGGAGAGTCC 808
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
 Db 809 CTGAGGGCAGTCAAGATCGCTGAGAAAGCGCAGAGAGGAAGATTCTGCCGAACCT 868
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 Db 869 ---GGAGTGTCCAAATAGGCAGTGGCCGAGCAGCAGACCTCTCTGTAGCAGAGCCT 925
 QY 220 GluAsnAspGlnGlyValValSerGluAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 Db 926 GAGCTAGTCAGGGACCAAGTGTCCAGTCAGACAGACAGATGATAGATAATTCAGTTGAA 985
 QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyLysArgLeuGlyArgProLysPro 259
 Db 986 GGGAGGGGACTCAGGANAATACAGAGAAACCCCGAGGAAGCGGCAACCAAGCCT 1045
 QY 260 GluCysGluGlyTyArgProAsnAlaLeuTyrcysIleCysArgGlnProHisAsnAsn 279
 Db 1046 GAGTGTGAGGTTTACGACCCCAATGCCATGCTGTACTGTCCATCTGCGCCAGCCTCAACAAC 1105
 QY 280 ArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIle 299
 Db 1106 AGGTTTATGATCTGCTGTGATCGGTGTGAGAGGTGGTTCCATGGTGAATCTGTGGTATT 1165
 QY 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCys 319
 Db 1166 TCTGAGGCCCGAGGCGGCTCTCTGGAAGGAACGGGGAAGACTACATCTGCCCAAAATTGC 1225
 QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLys 339
 Db 1226 ACCATTTCGAAGTGCAGGATGAGCAACCGTAGCCACCGATGACAGACTCTCGG 1285
 QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrIleGlyThrIleGluGlnLys 359
 Db 1286 TGCAGATCTGTGGGTGCTGATGGCAGACACTGCACACAGCATAGGACATAGAGCAGAAG 1345
 QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGlyLysAlaAlaAsnProSerGlyLys 379
 Db 1346 TCCGAGAGAGCAGGCGCATAAAGGGTAGGATTGAGAGGAGGAGCAACCCCGCGGCAAG 1405
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 Db 1406 AAAAACTCAAGATATTCCAGCTGTCTGTAGAGGCTCTGTGTCTCTTAATGATTCGC 1465

QY 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 419
 DB 1466 CTTGGTGTTCAGTCTAGCACAGCTGACTCTGTGTATTGAGTAATGACTGCAATCTC 1525
 QY 420 LysHisAlaAlaThrMetLysPheLeuSerSerGlyGluGlnLysProLysPro 439
 DB 1526 AACACGCGAGCTACCAAGATTCTAAGTTCAGGTAAAGAACAAAAACAAACCC 1585
 QY 440 LysGluLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
 DB 1586 AGGARAAGGTCAAGACGAGCAGAAAGATTCTCCAAATGCAGTGTTCAGGTG 1645
 QY 460 GlyIleGlySerSerValHisLysArgProAlaProGluLysGluThrVal 479
 DB 1646 GCGATTAAATCTCTCTGTCACAGAGACTAGCGTCAGAGAAAGGAGAAACCCAGT 1705
 QY 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGlu 499
 DB 1706 AAGAAA---GTGATGTGGCTTCAGAGGTGAGACTTCTGGGAAGGAGCGCTGTGAG 1762
 QY 500 SerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThr 519
 DB 1763 AGCAGCACCACTCTCTGGGAGTGCACCACTACAAATGCTGTGAAGCCAGAGAGCCA 1822
 QY 520 AlaAlaPro-----SerProSerLeuLeuTyrLysCysMet 531
 DB 1823 GAGAGCCCACTGCACTCTGCCCACTTCCCTATTGAGTAATCCATG 1867
 RESULT 13
 AB002331
 LOCUS Human mRNA for KIAA0333 gene, partial cds. 6692 bp mRNA linear PRI 06-OCT-2001
 DEFINITION Human mRNA for KIAA0333 gene, partial cds.
 ACCESSION AB002331
 VERSION AB002331.1 GI:2224606
 KEYWORDS KIAA0333.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 Nagase, T., Ishikawa, K., Nakajima, D., Chira, M., Seki, N., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
 JOURNAL DNA Res. 4 (2), 141-150 (1997)
 MEDLINE 94343984
 PUBMED 9205841
 REFERENCE
 2 (bases 1 to 6692)
 Ohara, O., Nagase, T., Kikuno, R. and Nomura, N. Direct Submission
 TITLE Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 2532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 JOURNAL (E-mail: cdmainfo@kazusa.or.jp, Tel: +81-438-52-3913)
 FEATURES
 Location/Qualifiers
 1..6692
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HG0988"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pBluescriptII SK plus"
 1..6692
 /gene="KIAA0333"
 <1..2978
 /gene="KIAA0333"
 /codon_start=3
 /protein_id="BA20791.1"
 /db_xref="GI:2224607"
 /translation="ESKLEKAAQDINDSEPGDLGPKPKCEGYDNALYICIRQPHN
 NRTWICDRCSEWFHGDVGISEARGILLERNEDYICNPCTILOVDTHSETADQQ

ORIGIN

Alignment Scores:
 Pred. No.: 3,35e-58 Length: 6692
 Score: 1575.00 Matches: 295
 Percent Similarity: 89.12% Conservative: 0
 Best Local Similarity: 89.12% Mismatches: 0
 Query Match: 52.69% Indels: 36
 DB: 9 Gaps: 1
 US-09-787-016A-3 (1-562) x AB002331 (1-6692)
 QY 235 GluSerLysLeuGluGlyLysAlaAlaGlnAspLysLeuGluGluProGlyAspLeu 254
 DB 3 GAGAGTAAGTTGAGGAGAAAGCGCTCAGGACATCAAGATGAGGAGCGCTGGAGACTTG 62
 QY 255 GlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArg 274
 DB 63 GCGCCGACCGAGCGCTGAATGTGAGGCTTACGAGCCCAAGCGCTGTATTGCATTTGCCGC 122
 QY 275 GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluThrPheHisGly 294
 DB 123 CAGCTTCACAAACAGAGTTTATGATTCTGTGACCGCTGTGAGAAATGGTTTTCATGCC 182
 QY 295 AspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr 314
 DB 183 GATTGTGTGGCAATTTCTGAGCTCGAGGAGGCTTTTGAAAGGAATGGGAGACTAT 242
 QY 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
 DB 243 ATCTGCCCAACTGCACCACTTCTGCAAGTGCAGATGAGACTCATTCAGAAACCGCAGAT 302
 QY 335 GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
 DB 303 CAGCGAAGCTTAATGGAGACTTGAGATGCTGTGCGACCGATTGTACAGTATAGGA 362
 QY 355 ThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAla 374
 DB 363 ACAATAGAGCAGAGTCTAGCGAAGACCAAGGAGTAAGAGGTAGATTGAGAAGCTGCA 422
 QY 375 AsnProSerGlyLysLysLysLeuLysIlePheGlnPro----- 387
 DB 423 AATCCAAAGTGGCAGAAAGAACTCAAGATCTTCCAGCTGTGTCCCGGGCTGTGCCACC 482
 QY 387 ----- 387
 DB 483 CAGTCCCTGCTCTCTGCGAGGTAATTTGAAATGTGTGTGTAGAGCATCTCTGCTTTC 542
 QY 388 -----ValIleGluAlaProGlyAlaSerLysCysIle 398
 DB 543 ACACCTTTCGACTGCAATTAGCTGCAAGGTGATAGAGCGCTGTGTCCCTCAAAATGATT 602
 QY 399 GlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIle 418
 DB 603 GCGCCCGGTGTGTGTACGTGGCGAGCCCGACTCGTGTACTGTCAGTATGACTGTATC 662
 QY 419 LeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 438

663	DB	CTCAAAACACCCGAGCGCAAAATGAAGTTCTTAAGCTTCAGGTAAGAAACAGAGGCCAAAG	722
439	QY	ProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGluValaGln	458
723	DB	CCTAAAGAAAAGATGAAGATCAAGACGAGAGCCAGCTTCCGAATATGGCGTCTCAG	782
459	QY	AlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysLysIleThr	478
783	DB	GCAGGTATTAATAAATCTCTCTGTGTGCACAGAGAGACCAGCTCCAGAAAAAAGAGACCA	842
479	QY	VallLysLysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCys	498
843	DB	GTEAAGAAAGGCAGTGTGTGTGCTCTCCGCGGAGTGAAGCACTGGGAGAGAGACGCTGT	902
499	QY	GluSerSerThrProSerTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLys	518
903	DB	GAGAGCAGCAGCCGCTGTGTGGCGAGCGATCACAATTTACAATGCAGTTAAGCCAGAAAG	962
519	QY	ThrAlaAlaProSerProSerLeuLeuTyLys	529
963	DB	ACTGTGTGCTCTCGCGGTCACTCTGTATAAA	995

```

RESULT 14
AC108337
LOCUS       AC108337               linear   HTG 21-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-115I13, *** SEQUENCING IN PROGRESS
            ****, 2 unordered pieces.
AC108337
FEATURES             (1)
     source          Rattus norvegicus (Norway rat)
ACCESSION      AC108337
VERSION        AC108337.4  GI:23266147
KEYWORDS       HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE         Rattus norvegicus (Norway rat)

```

ORIGIN: Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 272404)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Aquilano, D., Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Balwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Fails, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuehwa, L., Loulsegod, H., Losado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Kapus, P., Martin, K., Martin, R., Martinez, E., Maloney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Satek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villacana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,R., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 272404)
Worley,K.C.

Direct Submission
Submitted (27-Jan-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 272404)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (21-Sep-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:21737648.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.

----- Genome Center

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GPMX
Center clone name: CH230-115113
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 238513 bases at least Q40
Consensus quality: 241910 bases at least Q20
Consensus quality: 244206 bases at least Q20
Estimated insert size: 262154; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 270622: contig of 270622 bp in length
* 270623 270722: gap of unknown length
* 270723 272404: contig of 1682 bp in length.
Location/Qualifiers
1. 272404
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-115113"
1. 1256
/note="wgs end extension
FEATURES
source
misc_feature

```

```
FEATURES
source
misc_fc
```


misc_feature
clone_end:Sp6
complement(1814..2663)
/note="clone_boundary"
clone_end:Sp6
site:EcoRI
end_sequence:BH294125"
62809..64046
/note="wgs_contig"
267959..268006
/note="clone_boundary"
clone_end:r7
site:EcoRI
end_sequence:BH294091"

ORIGIN

Alignment Scores:
Pred. No.: 6, 06e-48 Length: 272404
Score: 1372.50 Matches: 312
Percent Similarity: 44.34% Conservative: 25
Best Local Similarity: 41.05% Mismatches: 55
Query Match: 45.92% Indels: 368
DB: 2 Gaps: 7

US-09-787-016A-3 (1-562) x AC108337 (1-272404)

Qy 1 MetAspAspGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 128283 ATGATGATTAAGGGGAGCTCGAGCAATGAGGAGCACCAGGCTATCAAAACCCACCAGT 128342
Qy 21 LysGluPheArgGlyThrTGGlyPheArgGlyThrIleAlaLysArgGlyVala 40
Db 128343 AAGAGTTTCAGAAACCTGGGGTTTTCGAAGAACCAATGCCAACGAGGGTGCA 128402
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 128403 GGAGACACCGAGGTGGAGCCCGAGTGAGCAGCAACCA-----CAACAGCATAACCTCTCC 128456
Qy 61 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 128457 CTGGCGCGCAGTGGACGGCAGCCAAAGCGGACCGAAAGGGTAGAAGAGTTCCTTACTACT 128516
Qy 81 AlaArgArgGlyArgGlnProLysArgMetProValSerLeuGluAspSerGlyGluProThr 100
Db 128517 GTTCGGCGCGCAGGAGGAGAAATGTCGGGTGTCCTCGGAGGATTCAGTGAGCCCAAG 128576
Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 128577 TCCTCCACAGTCACTGATGTGGACACAGCTTCTGAGGGGAGCGTTGAAGCAGTCTTGAG 128636
Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 128637 ATCAGAGTGGCCCT---GCACCTGGCACCTTAGGAAAGAACATCTCTCTCTCTGAA 128693
Qy 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 128694 AAGGTGAAGAGGCTGATGAGAGAGAGACACCTCTGACAGTGCACAGTGTGTCTTACG 128753
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnProThrGluArgPro 180
Db 128754 CTGAAGGAACCTTCAGAACCGCTTCGAAGAAAGGAGCAGGAGAACCTGTGGAGAGTCC 128813
Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGlyProAlaGluThr 200
Db 128814 CTGAGAGSCATTCAAGATCGACTCGGAAGCAGCGCCGAGAGGAGATCCTGCCGAAGCT 128873
Qy 201 -----ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLys 216
Db 128874 GGGAGTGTCCAGATAGGAGGATTTGAGGAGGAGCACACCTCGC-----TGTAAG 128921
Qy 217 GlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspGluSer 236
Db 128922 CAGGAGCTGAGGCTAGCCAGGGACCCAGTGTCTCCAGTCAGAGAGAGATGACAGAGAAAT 128981

Qy 237 LysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArg 256
Db 128982 CAGTTGGGGGGAAGCGACTCAGGGAGTAAAGGAGAAAGCCCGAGGAGCTGGCAA 129041
Qy 257 ProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysAlaGlnPro 276
Db 129042 CCAAGCCTGAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGCATCTGCCCGCAGCT 129101
Qy 277 HisAsnAsn----- 279
Db 129102 CACACAAACAGGTAGTGGAGGGAGGGAACTGTGTAGAACCAAGGATGGAGACAGCGA 129161
Qy 279 ----- 279
Db 129162 GAGGACTGACATTTGCTTTTATGTCAAAACAGCAGTTGGCTTATTGAGAAAAGAGCAATGA 129221
Qy 279 ----- 279
Db 129222 GTCTGAAAAAAGAAACAGAGATGCAATGCAATGCACTGCCCTAGAGCTGCCCTCTACCATG 129281
Qy 279 ----- 279
Db 129282 GCACCCACGTGTGTCCTCATCGACGGGATGAATGCACAGTGCCTCAGTCTCTCAAGCCAT 129341
Qy 279 ----- 279
Db 129342 GTCAGATGGACACTGTAGATAGCGGAGAACCAACTGTTGTCTCTACAGATCTTATGGAG 129401
Qy 279 ----- 279
Db 129402 CGTTAGAGGATCCACCCCTGTATGTCCTTTGAGTTGACTGCCAGTGTGTGATGGGGCATC 129461
Qy 279 ----- 279
Db 129462 TCTGTGGATTTGTAGAGGTGAAGTTTCTGTTGCTGTGTGGAGATACTCTAAAGCCCTCT 129521
Qy 279 ----- 279
Db 129522 AATTCACTCTTAGTGTCTAAACCAAGAGATGAAGGGCTGAGGATGTGGCTCTGTGATAG 129581
Qy 279 ----- 279
Db 129582 AATGCTTACTCAGCACACATGAAGCCCTGGGTGTGATGCCAGTACCTCATATAATGGT 129641
Qy 279 ----- 279
Db 129642 AGGTGGCATATGCTGTAACTCAGCTCTCAGTGTAGAGGAGGAGATCAAGAAATTCAGA 129701
Qy 279 ----- 279
Db 129702 TCATCTTTGGCTAGCTAGGCGAGTTTAAGACAGGAGAAACCCCTGTCTCTAAAAGTAAGTAA 129761
Qy 279 ----- 279
Db 129762 ATAGAGAAAAAGATGAGGGGCTGGAGAGATGCTGGTGGTTAAGAGCACTGACTGCTC 129821
Qy 279 ----- 279
Db 129822 TTCCAGAGGTCTGAGTTGATTTCCAGCAACCAACATGGTGGCTCACACCATCTGTAAAT 129881
Qy 279 ----- 279
Db 129882 GAGATGCGAGTCTCTCTTCTGTGTGTCTGTGAGACAGCTACAGTGTACTCATTAATTA 129941
Qy 279 ----- 279
Db 129942 TTATTAAATAAGTAATACAGTGTATACATTAAGTCTTTTTTAAATAAGAAAAAGAT 130001
Qy 279 ----- 279
Db 130002 GAAGCAGATTGGAGTGTCTTAACAAAGATGGGGGCACTTGGATGTTTTTAGCTCTG 130061
Qy 279 ----- 279

130062 ATCTCACTTTCCCTTTATTTGTCATGCTCTTCACTACTAGATGTTGAAGTGTGACCCACA 130121
 QY 280 -----ArgPheMetIleCysCysAspAr 287
 130122 TCATCTGTTGGACAGTGACACCATGATCTTCCATTTAGTTATGATCTGCTGACCG 130181
 Db
 287 GCYSGlulGluThrPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLe 307
 130182 GTGTGAAGAGTGTTCCATGGGAGTGGCTGGTGGGATATCTGAGGCCCGGAGCGGCTCTCT 130241
 Db
 307 uGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGluValGlnAspG 327
 130242 GGAGAGGAATGGGAGGAGTACTACATCTGCCAAATTTGCACATTTTGCAGTGCAGGATGA 130301
 Db
 327 uThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyrArgProGlyAspAlaAspG 347
 130302 GACGAGTAGTACCGCGCGATGAGCAGGATGCTGGGTGCGAGACCTGAGGTGCTGATGG 130361
 Db
 347 YThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLY 367
 130362 CACAGACTGTACACCATAGGACAGTACAGAGTCCCGGAGAGACCGAGGCGCATAAA 130421
 Db
 367 sGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLys 387
 130422 CGGTAGGATTGAGAAGCGCGCAACCCCGAGCGCGCAAGAAAGCTCAAGATATTCCAGCC 130481
 Db
 387 oVallIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAla 406
 130482 CGTA-----AGTACCTGCTG---CCTTGGCAGGCGACATGTGTCA 130518
 Db
 RESULT 15
 AL732560/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP24-209P15 on chromosome 2, complete sequence.
 ACCESSION AL732560
 VERSION AL732560.11 GI:22474418
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 104663)
 Leongamornlert,D.
 Direct Submission
 Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Aug 23, 2002 this sequence version replaced gi:22204532.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-209P15 is from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

FEATURES

source location/Qualifiers
 1. 104663
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP24-209P15"
 /clone_lib="RPC1-24"

Alignment Scores:
 Pred. No.: 4,44e-47 Length: 104663
 Score: 1342.00 Matches: 422
 Percent Similarity: 22.93% Conservative: 32
 Best Local Similarity: 21.31% Mismatches: 95
 Query Match: 44.90% Indels: 1433
 DB: 10 Gaps: 10

ORIGIN

US-09-787-016A-3 (1-562) x AL732560 (1-104663)
 QY 1 MetAspArgLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 Db 51952 ATGATGTTAAAGGACCTGAGCCATGAGAGACCAAGGCTATCAACCCACCACT 51993
 QY 21 LysGluPheArgLysThrIleGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 Db 51892 AAGGAGTTCCAGAAAACTGGGGTTTTCGAAGAACCAAGATTGCCAACCTGAGGGTCA 51933
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 51932 GGAGACACGAGGTGGACCCAGTCAGCAGCAACCA-----CAGCAGCATAACTCTCC 51779
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 51778 CTGGCCGCGAGTGGAGCAACCAACAACTACTGAGAGGGTAGAAGAGTTCTTACCACG 51719
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 51718 GTTCGGCGCGAGGAAAAAGATGTGCGGTGTCTTGGAGGATTCAGTGAGCCCA 51659
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 51658 TCTTCCACAGTCACATGATGTGGAGACAGTTCGAGGGGAGCGTGTAAAGCAGTCTGAG 51599
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 51598 ATCAGAAGTGGCCCTGTATCTGACTCTCTTAGGG---AAAGAACAATCTGCTCTTCTGAA 51542
 QY 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 Db 51541 AAGCAAAAGAGGTGAAGAGGAAGAAGACCTTCTGACAGTGAAGTGTGCTTAAG 51482
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
 Db 51481 TTGAAGCAACTTCAGAACCGCTTCGGAGAAAGCGAGAGCAAGAACCTGTGGAGAGTCC 51422
 QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200
 Db 51421 CTGAGAGCGAGTCAAGATCGGCTGAGGAGAGCGCAGAGAGGAAGATTCTGCCAACT 51362
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 Db 51361 ---GGGAGTGTCCAAATAGGCAGTGGCCGAGCAGGACAGACCTCTCTGTAGCAGAGCCT 51305
 QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 Db 51304 GAGCTAGTACGGGACCAAGTGTCCAGTCAGAGACAGATGACATAGATAATCAGTTGGA 51245
 QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyLysArgProLysPro 259
 Db 51244 GGGAGAGCGATCTCAGGCGAATACAGAGCAAAACCCAGGGGAAGCGGCAAAACCAAGCCT 51185

Db	49024	GTCATTTGATGCTTTTATTTATTTATCCAGTCAGTATGATGATGACATTGCTGAATTTTAGTC	48965	Db	47944	TTTGTGTGCCGGCAGTCAAGGAAATAGGTTTGTATAGTTTCACAAAGATAGGCTTAC	47885
Qy	387	-----	387	Qy	387	-----	387
Db	48964	TGTGTAACCTTAGATTGGTAGAAGAAAGCAACTATCTTTAAATATTTATCTCGAAACT	48905	Db	47884	CATGTACCTGGGCTGACCTGAATTTTGGAAAGTCTCCTGCCCTTCCCTCCATGTGCTAG	47825
Qy	387	-----	387	Qy	387	-----	387
Db	48904	GGGAGCTCTCAAAATGTCAGACATATGTTCTGTTCAGTCTTCTCTGATATATCTTAGACTA	48845	Db	47824	AATGGTGTAAAGATTATGTCATCATGCACAGCTAACAAATGAAAAATGAATTTTATT	47765
Qy	387	-----	387	Qy	387	-----	387
Db	48844	AAOGGTCTCTAGTTTCTGAAATGATATCTCTGTCCTGTCCTTTTGGCAATCTCCCTCCAAA	48785	Db	47764	CATATAAACTATTATTAGAGCCATCCATAGTGAATGATACACCTTGCACCTTTTAGTCCCAG	47705
Qy	387	-----	387	Qy	387	-----	387
Db	48784	CCACTCAAAACACAAACAAACAAACAGGCTGCTTTCTTCAAAACTAAACAGTGGATCTC	48725	Db	47704	CACGTAGGAGGAGAAGCAGGCAAAATCTCTTGAGTTCAAGGCTGGCTGCTGCATATC	47645
Qy	387	-----	387	Qy	387	-----	387
Db	48724	TTCTCTTAGGGCAGGCACATTTACTCTTTTAAAGAAATTTCACTTGAAGGCTAGGGGT	48665	Db	47644	AAAGTTTCTGGCAGCAGAAATTTTCATAGTGAGACCCTCTCTCAAAAAACACCAATAAA	47585
Qy	387	-----	387	Qy	387	-----	387
Db	48664	GGTGGAGTCCAACTCTAAAGGCAGTATGATTCATATATATATATTTCCCTCTGAGGGC	48605	Db	47584	AATTCCTGTAGTGTCTTAATACCAGTGACCACCAATAGTGTCTCACTGAGCAGCTGATC	47525
Qy	387	-----	387	Qy	388	-----	388
Db	48604	TGGAGAGTGTCTGATTTCAAGGAACTAGGTTCAATTCATAGTGGCCATATGGTTCTC	48545	Db	47524	ACTAGTGTGACACATGATGTCGAAGTTGCTGCTGTCTGTCTGAGGCTCGTAGAGGC	47465
Qy	387	-----	387	Qy	391	-----	391
Db	48544	ATCGTCATTTATAGTTCAAGTGTGAGGAGTCTGATATCTCTGTCATGACAGACAGA	48485	Db	47464	TCCTGTGTCTCTTAAATGATTTGGCTGCTGGTGTTCAGTGTAGCAGCAGCTGCTGT	47405
Qy	387	-----	387	Qy	411	-----	411
Db	48484	CATGAAGCAAGACACACATACATATAAAGAAATTCOAATTTTGTGACAGTTATAAGAG	48425	Db	47404	GTATTGCAAGTAACTGCTGCTCAACACAGCAGCAGCTACCATGAGATTCTTAAGTTC	47345
Qy	387	-----	387	Qy	431	-----	431
Db	48424	GAAATTTTACCTGAGCTCATAGATGCTCATTCATGCAATAATAGCATGGCCCATCAAACTG	48365	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48364	ATTACCTGAGTTTAACTCTTCAGATCTCATGTTGTAAGGAGAACTGACTCTCTGAAG	48305	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48304	TTGTCTGACTCTGCTGCTACCGCTATTTCTATGTGCAAGCTGCGGTATATGCAATGC	48245	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48244	ATGCACACAAATATGTAACAGAAATCTTAATGAAATTTGTAACTCATTAAAGTCTACAT	48185	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48184	TTTTTGGCTCTTTTTCATAGTGGTTTATTTTAAATGATGCTGTCTAGGGAGTAGTG	48125	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48124	TGCAACAGTATAGATGCCCTCAGAGGACAGGCTATCTGACATCAGTTCTAAACACTAGA	48065	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48064	CTGCTCTTCTGTAAGCATCCTCTGAACCATGTGTTTCAGGCTCTCTTACTCTTACCT	48005	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431
Db	48004	CATTGGGACCTGCACATGTTTAGAGGCAATAGATCTGAAGCTCCCATGTGTGCAGAG	47945	Qy	431	-----	431
Qy	387	-----	387	Qy	431	-----	431

```

QY 457 ----- 457
Db 46805 CCACTCAGCTGGGATTACAGTCTTGTGCCCAAGCCTAAGTGCATAGTCTTTTATTTT 46746
QY 457 ----- 457
Db 46745 TAGGACTGTTATTTTGGATCTTCAAACTCCATAAGTAGGAAGCAGATAGCAGCTTGTTA 46686
QY 457 ----- 457
Db 46685 GGTATTACAGACCTTGTGCCCCAGCAGCCCCCTTCACCTTTGTTTTCAGCTTGTCTGACT 46626
QY 457 ----- 457
Db 46625 GCCTTGGCGGAAGCGGCTCATGCTTTTATCTTGGCTGGTTCAAGCTTTTATCTTGGCTGG 46566
QY 457 ----- 457
Db 46565 TTCAATGCTCAGCTGGCTTATAGACAGATAGAGCCTTCCCAATTACCAGGTGTTAAAAA 46506
QY 457 ----- 457
Db 46505 GAATGCTAAGGGTCTTTTGANGTGTGGCTAGCCATAAAAAAGATCAGATGTTCTACC 46446
QY 457 ----- 457
Db 46445 TAGCTTGGTTTGAAGGCTCTTCAGGTTCTAAATAAAAGGTGTATATGTCCTTTAATTGGT 46386
QY 458 -----Glnal 459
Db 46385 TGTGAAGCAAAAGTGTCACTTCAAACCTTAATAATTGGGATTTCACTGATTTATCCAGGT 46326
QY 459 aGlyIleLysIleSerSerValHisLysArgProAlaProGluLysGluThrThrVa 479
Db 46325 GGGGATTAAATCTCTCTGTGCACAGAGACTAGCGTCAGAGAAAGGGAACCCAGT 46266
QY 479 lLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysG1 499
Db 46265 GAAGAAA--GTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAAGGAGCGCTGTGA 46209
QY 499 uSerSerThrProSerTipAlaSerRepHisAsnTyrAsnAlaValLysProGluLysTh 519
Db 46208 GAGCAGCACCACTCCTGGGCAAGTGCACAACTACATGCTGTGAGCCAGAGAGCC 46149
QY 519 rAlaAlaPro-----SerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyVa 536
Db 46148 AGAGAAAGCCCACTGCACTCTGCCCCACCCCTATTGAGTAAATGTACGTATCACCCAAAGGC 46089
QY 536 lGlyLeuLeuAspProSerArgSerPheTIpIleAlaIleProTIpAlaCysProGly 555
Db 46088 TGGCTTTCCAGGCCCTCCCATC--ATCTGGGTGGCT--GCCTGGGGCTGTCTAGGA 46036

```

Search completed: May 1, 2004, 20:13:40
Job time : 5512 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 17:35:11 ; Search time 592 Seconds
(without alignments)
4294.234 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKGPDSNEAPKAKPTS.....RSFWIAIPWACPGISVAALC 562

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09787016/runat_30042004_083703_5023/app_query.fasta_1.711
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-THR_MIN=0 -ALIGN=15 -MODES=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09787016 -CGEN 1 1 333 -runat 30042004_083703_5023
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
------------	-------	--------------	----	-------------

RESULT 1
US-09-925-301-327
; Sequence 327, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05862
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ IDS: 1694
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1398)

ALIGNMENTS

1	1706	57.1	1764	9	US-09-925-301-327	Sequence 327, App
2	1092	36.5	645	15	US-10-029-386-25091	Sequence 25091, A
3	807	27.0	478	10	US-09-918-985-19127	Sequence 19127, A
4	255	8.5	536	15	US-10-029-386-11381	Sequence 11381, A
5	232	7.8	5975	10	US-09-873-367C-112	Sequence 112, App
6	232	7.8	6256	15	US-10-037-270-369	Sequence 269, App
7	232	7.8	6256	16	US-10-117-722-369	Sequence 269, App
8	232	7.8	6948	10	US-09-930-213-384	Sequence 315, App
9	214.5	7.2	2320	15	US-10-175-523-142	Sequence 142, App
10	203	6.8	315	9	US-09-764-864-471	Sequence 471, App
11	182	6.1	8346	13	US-10-263-929-20	Sequence 20, App
12	181	6.1	1404	15	US-10-427-590-31	Sequence 31, App
13	177	5.9	3787	10	US-09-930-213-384	Sequence 284, App
14	176.5	5.9	7869	9	US-09-954-456-1921	Sequence 1921, App
15	176	5.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
16	174	5.8	3609	9	US-09-917-800A-1715	Sequence 1715, Ap
17	171.5	5.7	2502	13	US-10-342-887-1080	Sequence 1080, Ap
18	171.5	5.7	2502	13	US-10-172-118-1080	Sequence 1080, Ap
19	170.5	5.7	3183	13	US-10-263-929-50	Sequence 50, Appl
20	170.5	5.7	4056	16	US-10-144-194A-85	Sequence 85, Appl
21	170	5.7	3664	16	US-10-144-194A-73	Sequence 73, Appl
22	169.5	5.7	5225	15	US-10-253-157-109	Sequence 109, App
23	169	5.7	2031	10	US-09-823-187-23	Sequence 23, Appl
24	169	5.7	2031	10	US-09-863-776-13	Sequence 13, Appl
25	169	5.7	2031	15	US-10-203-708-5	Sequence 5, Appl
26	169	5.7	2154	15	US-10-203-708-6	Sequence 6, Appl
27	169	5.7	2390	15	US-10-102-524-1739	Sequence 1739, Ap
28	168	5.6	2382	15	US-10-084-817-312	Sequence 312, App
29	168	5.6	3174	13	US-10-282-122A-30251	Sequence 30251, A
30	167	5.6	2004	10	US-09-863-776-15	Sequence 15, Appl
31	167	5.6	2004	10	US-09-863-776-17	Sequence 17, Appl
32	167	5.6	4530	13	US-10-302-172-881	Sequence 881, App
33	166.5	5.6	3727	15	US-10-171-581-49	Sequence 49, Appl
34	166	5.6	1938	15	US-10-106-698-1570	Sequence 1570, Ap
35	166	5.6	2272	10	US-09-946-374-307	Sequence 307, App
36	166	5.6	2272	12	US-10-015-395A-307	Sequence 307, App
37	166	5.6	2272	13	US-10-206-915-345	Sequence 345, App
38	166	5.6	2272	13	US-10-199-670-345	Sequence 345, App
39	166	5.6	2272	13	US-10-201-858-345	Sequence 345, App
40	166	5.6	2272	13	US-10-205-890-345	Sequence 345, App
41	166	5.6	2272	13	US-10-208-024-345	Sequence 345, App
42	166	5.6	2272	13	US-10-201-853-345	Sequence 345, App
43	166	5.6	2272	13	US-10-174-581-345	Sequence 345, App
44	166	5.6	2272	13	US-10-176-483-345	Sequence 345, App
45	166	5.6	2272	13	US-10-176-749-345	Sequence 345, App

OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1758)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1759)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (1762)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-925-301-327

Alignment Scores:
 Pred. No.: 5,618-148 Length: 1764
 Score: 1706.00 Matches: 317
 Percent Similarity: 99.37% Conservative: 0
 Best Local Similarity: 99.37% Mismatches: 1
 Query Match: 57.08% Indels: 2
 DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x US-09-925-301-327 (1-1764)

Qy 245 AspileLysAspGluProGlyAsp-LeuGlyArgProLysProGluCysGluGlyTy 264
 Db 2 GACATCAAGATGAGCGCTGAGACTTTGGCGGACCGAAGCTGGAATGAGGGTTA 61
 Qy 264 rAspProhnaLalaLeuTyrcyleCysArgGlnProhnaLalaLeuTyrcyleCys 284
 Db 62 CGACCCCAACGCCCTGATTTGATTTGGCGGAGCTTCAACACAGGTTTATGATTG 121
 Qy 284 cYsAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgG 304
 Db 122 CTGTGACCGCTGTGAGATGTTTCATGGCGATTTGTGGCATTTCTGAGGCTCGAGG 181
 Qy 304 YArgLeuLeuGluArgAsnGlyGluAspTyrcyleCysProhnaLalaLeuGluGlnVa 324
 Db 182 GAGGCTTTTGGAAAGGAATGGGAAAGACTATATCTGCCCAACTCACCATTCTGCAAGT 241
 Qy 324 GlnAspGluTrpHisSerGluThraAlaAspGlnGluAlaLysTrpArgProGlyAs 344
 Db 242 GCGGATGAGACTCATTCAGAACGGCGATCAGCAGAGAGCTTAATGGAGACCTGGAGA 301
 Qy 344 pAlaAspGlyThrAspCysThrSerileGlyThrleGluGlnLysSerSerGluAspG 364
 Db 302 TGCTGATGCGACCGATTGTACAGTATAGCAATAGCAGCAGAGTCTAGCGAAGACCA 361
 Qy 364 nGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLys 384
 Db 362 AGGGATAAGGGTAGATTGAGAAAGCTGCAATCCAGTGGCAAGAGAACTCAGAT 421
 Qy 384 ePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysEHi 404
 Db 422 CTTCAGAGCTGTGATAGAGCGCTGTGCTCAAAATGATTGGCCCGGGGTGCTGCA 481
 Qy 404 sValAlaGlnProAspSerValTyrcysSerAsnAspCysIleLeuLysHisAlaAla 424
 Db 482 CGTGGCGCA-CCGACTCGGTGTACTGCGAGTAAATGACTGTATCTCTCAACACGCGCGAGC 540
 Qy 424 aThrMetLysPheLeuSerSerGlyLysGlnLysProLysProLysGluLysMetLy 444
 Db 541 GACATGAAGTTCTAAGCTCAGGTAAAGACAGAGCCCAAGCCCTAAAGAAAGATGAA 600
 Qy 444 sMetLysProGluLysProSerleupProLysCysGlyAlaGlnAlaGlyIleLysIleSe 464
 Db 601 GATGAGCCGAGAGAGCCCGATTTCCGAAATGCGGTGCTCAGCGAGGTTTAAATCTC 660
 Qy 464 rSerValHisLysArgProAlaProGluLysGluThrThrValLysLysAlaValVa 484
 Db 661 TTTCTGTGCACAGAGACCGAGCTCCAGAAAAAAGAGACCAAGTGAAGAGGCGATGTT 720
 Qy 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThProSe 504
 Db 721 GGTCTCCGCGAGTGAAGCACTCGGAAGGAAGCAGCTTCTGAGAGCAGCAGCCCTC 780

Qy 504 rTPAlaSerAspHisAsnTyrcysAsnAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 781 GTGGCGGAGCGATCACAATTACAAATGAGTAAGAGCAGAAAGAGACTGCTGCTCCTCGCC 840
 Qy 524 oSerLeuLeuTyrcysCysMetTyrcysHisLeuGlyValGlyLeuLeuAspProSerArgSe 544
 Db 841 GTCACTGTTGTATTAATGTATGATATACCTACCCAGGGTTGGCTCTCTGAGCCCTCCGTTTC 900
 Qy 544 rPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAlaLeuCys 562
 Db 901 TTTCTGGATAGCCATCCCTTGGGCTGTCTCAGGACTGGGAGTTGCGAGTTTGTGT 955

RESULT 2

US-10-029-386-25091/c

; Sequence 25091, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AROMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Arnonax Sequence Listing Engine vers. 1.1

; SEQ ID NO 25091

; LENGTH: 645

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO ALL17379,14

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1

; OTHER INFORMATION: SWISSPROT HIT: P23583, EVALUATE 2.50e-01

; OTHER INFORMATION: EST_HUMAN HIT: AL580669.1, EVALUATE 0.00e+00

; OTHER INFORMATION: NT HIT: G116193858, EVALUATE 0.00e+00

US-10-029-386-25091

Alignment Scores:

Pred. No.: 9e-92 Length: 645

Score: 1092.00 Matches: 214

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 36.53% Indels: 0

DB: 15 Gaps: 0

US-09-787-016A-3 (1-562) x US-10-029-386-25091 (1-645)

Qy 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
 Db 643 ATGACCGACCAAGCGACCGCGAGCAATGAGGAGCCTTAAGGCCATCAAAACCCACCGAGC 584
 Qy 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
 Db 583 AAAGATTTCAGAAAAAATGGGGTTTTCGAGGACCACTATCGCCAGAGAGAGGGCGCA 524
 Qy 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 523 GGGGACGCGGAGGCTGACCCACTGAGCGCGCCACCCACACAGCAGCAGCTGCGGCTGTGCC 464
 Qy 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 Db 463 CTGGCGCGCGTGGGAGGAGCAGCCCAAGCGGCTGAGCGCGTGGAGGAGTCTCTGACCATTT 404
 Qy 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 Db 403 GCGCGGCGCGCGGAGGAGCATGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 344
 Qy 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 343 TCTGCCCCCGCCACAGACGCGCGGAGCAGCTCTCCAGGGGCGAGCGTGGAAAGCGCTTCTGAG 284

Db 286 GATTGTGTGGCATTCTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAGACTAT 345
Qy 315 IICQYCPKASNCYETHRILELEUINVALGINASPLNTHRHISERGLUThrAlaasp 334
Db 346 ATTCGCCCAACCTGCACTTCTGCAAGTGCAGATGAGACTCATTCAGAAAAGGAGAT 405
Qy 335 GlnGlnGluAlaLysPArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
Db 406 CAGCAGGAGCAATAATGAGAGCTCGAGATGCTGATGGCACCAGGATTTGTACAGTATAGGA 465
Qy 355 ThrIleGluGln 358
Db 466 ACAATAGAGCAG 477
RESULT 4
US-10-029-386-11381/c
; Sequence 11381, Application US/10029386
; Publication No. US20030134704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Shazron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ABOVICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11381
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL17379.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P80398, EVALUE 3.80e+00
; OTHER INFORMATION: NT HIT: G14769835, EVALUE 1.00e-78
; OTHER INFORMATION: EST_HUMAN HIT: BF205331.1, EVALUE 2.00e-78
US-10-029-386-11381
Alignment Scores:
Pred. No.: 1.54e-14 Length: 536
Score: 255.00 Matches: 48
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.53% Indels: 0
DB: 15 Gaps: 0
US-09-787-016A-3 (1-562) x US-10-029-386-11381 (1-536)
Qy 1 MetaAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
Db 146 ATGAGCAGCAAAAGCGGACCCGAGCATGAGAGGACCTTAAGGCCATCAACCCACCCAGC 87
Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 86 AAAGAGTTTCAGAAAACATGGGTTTTCGAAGGACCACTATGCGCAAGCGGAGGGGCA 27
Qy 41 GlyAspAlaGluAlaAspProLeu 48
Db 26 GGGGACGCGGAGGCTGACCCACTG 3
RESULT 5
US-873-367C-112
; Sequence 112, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena

Qy 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 283 ACCAGAGCGGCCCCCTCTGCTTCACAGCTGTGAGGAGCAACAGCAGCTCTTCGAA 224
Qy 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 223 AAGGTGAAGAGGAGGGATGACACCATGACCTCCGATAGTACAGCGATGGCTGACC 164
Qy 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGlnGlnGluProThrGluArgPro 180
Db 163 TTGAAGAGAGCTTCAGAAATCGCTTCGAGGAGCGGGGAACAGGAGGCCACTGAGAGGCC 104
Qy 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGlnGluGlyProAlaGluThr 200
Db 103 CTGAAGAGGATCCAGAGTGGCTCGCGAGAGCGCGGGAGGAGGGTCCCGCGGAGACT 44
Qy 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro 214
Db 43 GTGGGCTCCGAGGCGAGTGACACTGTGGAGGCGCTCTCTGCC 2
RESULT 3
US-09-918-995-19127
; Sequence 19127, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19127
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(478)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19127
Alignment Scores:
Pred. No.: 1.35e-65 Length: 478
Score: 807.00 Matches: 144
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.00% Indels: 0
DB: 10 Gaps: 0
US-09-787-016A-3 (1-562) x US-09-918-995-19127 (1-478)
Qy 215 SerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArg 234
Db 46 AGTAAGCAGGAGCGGAGAGCAGATCAGGGGGTGTGTCCTCCAGGCTGGGAAGATGACAGA 105
Qy 235 GluSerLysLeuGluGlyLysAlaGlnAspLysAspGluGluProGlyAspLeu 254
Db 106 CAGAGTAAGTTGGAGGAAAGCGGCTCAGGACATCAAGATGAGGAGCCTGGAGACTTG 165
Qy 255 GlyArgProLysProGluCysGluGlyThrAspProAsnAlaLeuThrCysIleCysArg 274
Db 166 GGCGAGCGAAGCTCGAATGTGAGGGTTACACCCCAAGCCCTGTATGTGATTTGCCGC 225
Qy 275 GlnProHisAsnAspArgPheMetIleCysCysAspArgCysGluGluThrPheHisGly 294
Db 226 CAGCTCACACACAGAGTTATGATTTGCTGTGACCGCTGTGAAGATGTTTCATGGC 285
Qy 295 AspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr 314

```

1  APPLICANT: Ebner, Reinhard
2  APPLICANT: Carter, Kenneth
3  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
4  TITLE OF INVENTION: Signature Gene Sets
5  FILE REFERENCE: 689290-64
6  CURRENT APPLICATION NUMBER: US/09/873,367C
7  CURRENT FILING DATE: 2003-04-29
8  PRIOR APPLICATION NUMBER: U.S. 60/236,891
9  PRIOR FILING DATE: 2000-09-29
10 PRIOR APPLICATION NUMBER: U.S. 60/236,842
11 PRIOR FILING DATE: 2000-09-29
12 PRIOR APPLICATION NUMBER: U.S. 60/244,867
13 PRIOR FILING DATE: 2000-11-01
14 PRIOR APPLICATION NUMBER: U.S. 60/245,084
15 PRIOR FILING DATE: 2000-11-01
16 NUMBER OF SEQ ID NOS: 1067
17 SOFTWARE: Patent in version 3.0
18 SEQ ID NO 112
19 LENGTH: 5975
20 TYPE: DNA
21 ORGANISM: Homo sapiens
22 US-09-873-367C-112

```

Alignment Scores:			
Pred. No.:	3,67e-11	Length:	5975
Score:	232.00	Matches:	94
Percent Similarity:	36.43%	Conservative:	63
Best Local Similarity:	21.81%	Mismatches:	154
Query Match:	7.76%	Indels:	120
DB:	10	Gaps:	17
US-09-787-016A-3 (1-562) x US-09-873-367C-112 (1-5975)			
Qy	92	SerLeuGluAsp---SerGlyGluProThrSerCysProAlaThrAspAlaGluThrAla	110
Db	837	TCITTCAGTGATAAGTGCACAGCCTCATCTCGTGTGCTTGAAAGAACCCTCATCATCTGCCA	896
Qy	111	SerGluGlySerValGluSerAlaSerGluThrArg---SerGlyProGlnSerAlaSer	129
Db	897	CAAACTGGACATGATCATCATCTACGAAACAGTGTCAATAGCCTCAGCAACAGGCC	956
Qy	130	ThrAlaValIysGluArgProAlaSerSerGluIysValIys---GlyGlyAspAspHis	148
Db	957	CCAGCAATGAAACCAATAGTCACGTGAAGGAAGAGCTTGAACACCCAGGCGTTGAGCAT	1016
Qy	149	AspAspThrSerAspSerAspSerAspGlyLeuThrLeuIysGlu---	164
Db	1017	-----TTTAAGGAAGAGGTAAACTGAAACTGMAAAACCTGAGAAGAACCTA	1064
Qy	165	GlnAsnArgLeuArgArgIysArgGluGlnGluProThrGluArgProLeuIysGlyIle	184
Db	1065	CAACCCGCGCAAGAGAAGACGACAAAGTTTCTTTAGATGAGCCACCATTTGTTCAAT	1124
Qy	185	GlnSerArgLeuArgIysIysArgArgGluGluGlyProAlaGluThrValGlySerGlu	204
Db	1125	CCGATAAATCACTACCATACCATAGAGAGAA-----GGCTCT---	1160
Qy	205	AlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGly	224
Db	1160	-----	1160
Qy	225	ValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGln	244
Db	1161	-----GATCATAGCTCCTCATTTTGAAGCAAAATAT-----	1190
Qy	245	AspIleIysAspGluGluProGlyIysAspLeuGlyArgProLysProGluCysGluGlyTyr	264
Db	1191	-----ATGTGACTCCGACGACGAGTGT-----GGGTTT	1220
Qy	265	AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys	284
Db	1221	-----TGCAAAACCAACCATGGCAACAGGTTTATGTTGTCGC	1256

1586 CAGCAGATGGCGAGGAGCAACAAAGTAATGCTGTGTAATAATGTTGT-----GCT 1636
325 GlnAspGluThrHisSerGluThrAlaAsp-----GlnGlnGluAla 338
1637 GAAGAGACCAAAAGACTGAAATPACTAGATCCAGATCTTTGGAAACCAAGCTACAGTT 1696
339 LysTrpArgProGlyAspAlaAspGlyThrSerCysThrSerIleGlyThrIleGluGln 358
1697 GAATTCATAGTGGAGATAAACA---ATGGAGTGTGAAAAGCTTGATATTCAAACAC 1753
359 LysSerSerGlu-----AspGlnGlyIleLysGlyArgIleGluLysAla 373
1754 ACAACAATATAGTAACCAATATATAGATGATACAGTGAAGCAACAGGTCAAAATTTTA 1813
374 AlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGly 393
1814 AAACGGGAGTCTGCTGCAAGCAGA----- 1837
394 AlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyCys 413
1838 -----AATTCATCAGACTGT 1852
414 SerAsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLys 433
1853 AGAGATAATGAAATTAATAAATGCGCAGTAGCTCTCTGCTAAGATGGGACCAACGATT 1912
434 GluGlnLysProLysProLysGluLysMetLysMetLysProGlyLysProSerLeuPro 453
1913 TTACTCGGAGATCTCCAGAAAGAAAAGTGAATAAATACCGAAGAGCTTACAACTGTT 1972
454 LysCysGlyAlaGlnAlaGlyIleLysIleSerValHisLysArgProAlaProGlu 473
1973 ACTTGC-----ACAGGAGAAAGAGCTTCAAAACAGGTACTCATGAGAGCAGAG 2023
474 LysLysGluThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGly 493
2024 ATGAAAGAGAGAAAGTGTGAAAAGGAGTG-----CTTAAT 2059
494 LysGluAlaAlaCysGluSerSerThrProSer 504
2060 GTACATCTCTCTGCTCTGCTTCTCAAGCCTTCT 2092

RESULT 7
US-10-117-722-269
; Sequence 269, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radcoje T.
; TITLE OF INVENTION: No. US20030219744A1elel Nucleic Acids and
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: PCT FL Genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257) ... (5443)
US-10-117-722-269

Alignment Scores:
Pred. No.: 3,87e-11 Length: 6256
Score: 232.00 Matches: 94
Percent Similarity: 36.43% Conservative: 63
Best Local Similarity: 21.81% Mismatches: 154
Query Match: 7.76% Indels: 120
DB: 15 Gaps: 17

US-09-787-016A-3 (1-562) x US-10-037-270-269 (1-6256)

Qy 92 SerLeuGluAsp---SerGlyGluProThrSerCysProAlaThrAspAlaGluThrAla 110
Db 1106 TCTTTGATGATAGTACACAGCTATCTGCTGTTGAAAGAACCTCATCTCTGCA 1165
Qy 111 SerGluGlySerValGluSerAlaSerGluThrArg---SerGlyProGlnSerAlaSer 129
Db 1166 CAACCTGGACATGATACATCTTAGCCAGAACAGTGTCTAAGCCTCAGCAACAGGCC 1225
Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLys---GlyGlyAspAspHis 148
Db 1226 CCAGCAATGAAACCAATAGTCACTGAGGAGAGAGCTTGAACACCCAGCGGTGAGCAT 1285
Qy 149 AspAspThrSerAspSerAspGlyLeuThrLeuLysGlu-----Leu 164
Db 1286 -----TTTAGGAGAGAGTAACTGAACTGAAATAAACCCTGAGAGAACCTA 1333
Qy 165 GlnAsnArgLeuArgArgLysArgGlnGluProThrGluArgProLeuLysGlyIle 184
Db 1334 CAACCCCGCAAGAGAGAGCAAAAGTCTTTCTTAGATGAGCCACCATTTCTTCATT 1393
Qy 185 GlnSerArgLeuArgLysArgGluGluGlyProAlaGluThrValGlySerGlu 204
Db 1394 CCAGATACATAGTACCTATAGAGAGAA-----GGCTCT-- 1429
Qy 205 AlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGly 224
Db 1429 ----- 1429
Qy 225 ValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGln 244
Db 1430 -----GATCATGCTCTCTCAATTTGAAGCAATAT-- 1459
Qy 245 AspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTy 264
Db 1460 -----ATGTGGACTCCCAAGCAGCTGT---GGGT 1489
Qy 265 AspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 284
Db 1490 -----TGCAAAAACCAACATGCAACAGTTTATGTTGGC 1525
Qy 285 CysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGly 304
Db 1526 TGTGGAGATGTGATGACTGGTTTCATGCTGTTGTTGGTTAAGTCTTTCTCAAGCA 1585
Qy 305 ArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThrIleLeuGlnVal 324
Db ----- 324

1930 CCAGCAATGAAACCAACCAATAGTCACGTGAAGGAGAGCTTGAACACCCAGCCGCTTGAGCAT 1989
149 AspAspThrSerAspSerAspGlyLeuThrLeuLeuGlu-----Leu 164
1990 -----TTTAAAGGAAGAGGATAAATGAACCTGAAATAACCTGAGAAGAACCTA 2037
165 GlnAsnArgLeuArgArgGlyArgGluGlnGluProThrGluArgProLeuLeuGlyLeu 184
2038 CAACCCCGCCAAAGAAAGACGACCAAAAGTTTCTTTAGATGAGCCACCACTGTTTCATT 2097
185 GlnSerArgLeuArgGlyArgGlyArgGluGluGluProAlaGluThrValGlySerGlu 204
2098 CCGGATAACATAGTACCATAAGAGAGAA-----GGCTCT---- 2133
205 AlaSerAspThrValGluGluValLeuProSerLysGlnGluProGluAsnAspGlnGly 224
2133 ----- 2133
225 ValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGln 244
2134 -----GATCATAGTCTCTCATTTGAAAGCAAAATAT----- 2163
245 AspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyr 264
2164 -----ATGTGCACTCCACGACGAGTGT--GGGTTT 2193
265 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 284
2194 -----TGCAAAACCAACCATGCGCAACAGGTTTATGGTTGCG 2229
285 CysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAlaArgGly 304
2230 TGTGGGAGATGTCAGTCTGTTTCATGTTGATGTTGGTTGGTTAAAGTCTTTCTCAAGCA 2289
305 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 324
2290 CACGACATGGCGGAGGAAGCAAAAGAAATATGCTGTGTAATAATGTGT-----GCT 2340
325 GlnAspGluThrHisSerGlnThrAlaAsp-----GlnGlnGluAla 338
2341 GAGAAGACAAAGAACTGAATACTAGATCCAGATACITTTGGAACCAACCAAGTACAGTT 2400
339 LysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGln 358
2401 GAATTCATAGTGGAGATAAACA-----ATGGAGTGTGAAGAGCTTGATTTACAAACAC 2457
359 LysSerSerGlu-----AspGlnGlyIleLysGlyArgIleGluLysAla 373
2458 ACACCAATGATAGAACCAAAATATATAGATGATGATGATGATGATGATGATGATGATGAT 2517
374 AlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGly 393
2518 AAACGGAGTCTGTGAAGGCAG----- 2541
394 AlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyrCys 413
2542 -----AATTCATCAGACTGT 2556
414 SerAsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLys 433
2557 AGAGATAATGAATTAATAAATGCGAGTACTCTCTCTGTAAGATGGGCAACCAAGTT 2616
434 GluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSerLeuPro 453
2617 TTACTCTGGAGATCTCTCAGAGAAAGAAAGTGAATAATACCGAAAGAGTCTACACTGTT 2676
454 LysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGlu 473
2677 ACTTGC-----ACAGAGAGAAAGCTTCAAAACCAAGTACTCATGAGAGAGCAG 2727
474 LysLysGluThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGly 493

Db 2728 ATGAAAAGAGAGAGAGTTGAAAAGGAGTG-----CTTAAT 2763
Qy 494 LysGluAlaLysCysGluSerSerThrProSer 504
Db 2764 GTACATCTGCTGCTCTGCTTCCAAAGCCTTCT 2796
RESULT 9
US-10-175-523-142
; Sequence 142, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHPTS)
; FILE REFERENCE: 3235/13795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 142
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-175-523-142
Alignment Scores:
Pred. No.: 4,9e-10 Length: 2320
Score: 214.50 Matches: 95
Percent Similarity: 38.38% Conservative: 42
Best Local Similarity: 26.61% Mismatches: 98
Query Match: 7.18% Indels: 122
DB: 15 Gaps: 20
US-09-787-016a-3 (1-562) x US-10-175-523-142 (1-2320)
Qy 224 GlyValValSerGln---AlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAla 242
Db 24 GGGTCTGTGAGGAGTGCGCGGAGCGAGATATGAGGGA----- 65
Qy 243 AlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGlu 262
Db 66 -----GATGGTTCAGACCCAGAGCTTCAGATGCCGGGAGGAGCAGC---AAGTCCGAG 116
Qy 263 GlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMet 282
Db 117 AATGGGAGAAATGCGCCCATCTACTGTCATCTCCGCAACCGGACATCACTGCTTCATG 176
Qy 283 IleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAla 302
Db 177 ATCGGGTGTGACACTGCAATGAGTGTTCATGCGGACTGTCATCGGAGTCTGAGAAAG 236
Qy 303 ArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeu 322
Db 237 ATGCCCAAGCCATC-----CGGAGTGTGACTGTGCGGAGTGC----- 275
Qy 323 GlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArg---- 341


```

RESULT 10
US-09-764-864-471
; Sequence 471, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (894)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

RESULT 11

Db 194 ACCACTCGCAGCGCGCGCGCGCGCTGCTGGACACCGTGAACCTACCCCATCCACATGA 253
Qy 87 ArgSerMetProValSerLeuLysSerGlyGluProThrSerCysProAlaThrAsp 106
Db 254 AGAACCTGTCCC---TCAAGGAGCTGCAGCAGCTCTCCGACGAGCTGCGCTCCGACGTCA 310
Qy 107 AlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGln 126
Db 311 TCTTCCACGTCTCCAAAGACCGCGC-----GCACCTCGGGTCCAGCC 352
Qy 127 SerAlaSerThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyLys 146
Db 353 TCGGGTGTGCTGAGCTCACGTGCGCTGCTACTAGCTTCTCAACACCCCGC----- 403
Qy 147 AspHisAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsn 166
Db 404 -----AGGACAAAGCTCTCTCGGAGCGTCCGCCACCACTGCTACCCG-CACAAG 450
Qy 167 ArgLeuArgArgLysArgGluGluGluProThrGluArgProLeuLysGlyLysLeuSer 186
Db 451 ATTCTGACGGCGCGCGCGATAGATGCCGAGATGCGGACAGTCCGACACACCGCTCTCCGGC 510
Qy 187 ArgLeuArgLysArgGluGluGluProAlaGluThrValGlySerGluLysLeu 206
Db 511 TTGGTC-----NAGCGCTCCGAGAGCGAGTACGACAGCTTCGGCAGCGCGCACAGC 561
Qy 207 AspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyVal 226
Db 562 TCCACCACCATCTCCCGCGC-CCTCGGATGCGCGTCCGAGGAGCACTCAAGGGCGGAA 620
Qy 227 SerGlnAlaGlyLysAspAspArgLysLeuGluGlyLysAlaAlaGlnAspLe 246
Db 621 GAACAACGTGTGGGGTGTGGGACGGGCGCATGACGCGCGCGCAGGC-----GTA 674
Qy 247 LysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyAspPro 266
Db 675 CGAGCGATGAACAACGCGCGCTACTCGA-----CTCGACATGAT--- 716
Qy 267 AsnAlaLeuTyCysLysCysArgGlnProHisAsnAsnArgPheMetLysCysAsp 286
Db 717 -----CGTGATCTCTCAACGACCAACAGCAGGCTGCTGCGCGAC 755
Qy 287 ArgCysGluGluTrpPheHisGlyAspCysValGlyLysSerGluAlaArgGlyArgLeu 306
Db 756 -----GGCAGCGCTCGACGGCGCGCGCGCGCGCGCGCTGCGCGCGCT 794
Qy 307 LeuGluArgAsnGlyLysAspTyLysCysProAsnCysThrLeuGlnValGlnAsp 326
Db 795 CAGCGCGC----- 803
Qy 327 GluThrHisSerGluThrAlaAspGlnGlnAlaLysTrpArgProGlyAspAlaAsp 346
Db 804 -----CCTCAGCAAGCTGCTCAGCGCGCGCTCAGGAGCTGAGGAGGTGGCAA 857
Qy 347 GlyThrAspCysThrSerLysGlyThrLeuGlnLysSerSerGluAspGlnGlyLe 366
Db 858 GGGAGTGAAGCAAGCAAT-----CGCGGGTGGTGGCAGAGATCGCGCGCAA 905
Qy 367 LysGlyArgLysGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysLysPheGln 386
Db 906 GGTGACGAGTACGGCC-GCGCATGATCAGCG----- 937
Qy 387 ProValLysGluAlaProGlyAlaSerLysCysLysGlyProGlyCysHisValAla 406
Db 938 -----GCTCGGGTGTGCTCTTCGAGGAGCTCGGGCTGTATT---ACATCG 982
Qy 407 GlnProAspSerValTyCysSerAsnAspCysLysLeuLysHisAlaAlaThrMet 426
Db 983 GCCCGCTGACGGCC-----ACAACTGACGACC 1012
Qy 427 LysPheLeuSerSerGlyLysGluGlnLysProLysPro----- 439
Db 1013 TCATCACCATCTCTGGGAGGTCAAGGGCACCAGACCCCGCGCGCGTGTCTCATCCATG 1072

Qy 440 -----LysGluLysMetLysMetLysProGluLysProSerLeuProLysCys 455
Db 1073 TCATCACCAGAAAGCGCGGCTACCCCTACCGCGAGCGCTCCGACAAAGTACCAAC 1132
Qy 456 GlyAlaGlnAlaGlyLysLysSerSerValHisLysArgProAlaProGluLysLys 475
Db 1133 GGTGGCAAGTTCGATCCGCGACCGGAGGAGTTCAGGGTCCGGCCCAAGACGCTT 1192
Qy 476 GluThrThrValLysLysAlaValValProAlaArgSerGluAla-----Leu 492
Db 1193 CCTACACA-----ACTACTCGCGAGCGCTCATAGCCGAGCGG 1234
Qy 493 GlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerAspHisAsnTyAsn 512
Db 1235 GGCAGACACAGATGCTGCGCATTCACCGGCCATGGGGCGGAGACGGGCTCACT 1294
Qy 513 AlaValLysProGluLysThrAlaAlaProSerProSerLeuLeuTyLysCysMetTy 532
Db 1295 ACT-----TCTCCGCGCTTCCCAACC----- 1318
Qy 533 HisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpLeuAlaLeuProTrpAla 552
Db 1319 -----GGGCTTCCAGTCCGATCCGCT-AAACAGAACCGCTTACCCCTCCGCGCGC 1371
Qy 553 CysProGlyLeuGlyValAlaAlaLeuLys 562
Db 1372 CTGCGCGCAGGGGTAAACCTTCTGCTGC 1401
RESULT 13
US-09-930-213-284
; Sequence 284, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHÉ-NITSE, OLEG
; APPLICANT: GRIPS, MARTIN
; APPLICANT: HELLMER, MARTIN
; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SERS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBRS-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 284
; LENGTH: 3787
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-930-213-284
Alignment Scores:
Pred. No.: 2,56e-06 Length: 3787
Score: 177.00 Matches: 108
Percent Similarity: 36.18% Conservative: 49
Best Local Similarity: 24.88% Mismatches: 162
Query Match: 5.92% Indels: 119
DB: 10 Gaps: 20
US-09-787-016A-3 (1-562) x US-09-930-213-284 (1-3787)
Qy 189 ArgLysLysArgArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThr 208
Db 141 CGGCTCCCGCGCGCGCCAGCAGCGCGCTGATGTGGACGCGCGCGCGCGCGCG 200
Qy 209 ValGluGly---ValLeuProSerLysGlnGluProLysAsnAspGlnGlyValValSer 227

[illegible]

```

Qy 527 -----LeuTyrIlybCybMetTyrHisIleuG 535
Db 1223 ACAGCAGAGATGTTCTTTCTGCTGCACAGGTCGACAAATGCTACAAAGTCATGCTCAAGCAGG 1282
Qy 535 IyValGlyIleuLeuAppProSerArgSerPheIrpIle 547
Db 1283 GCCAGACCTCTTCATCCCTCAGGC-----TGGATC 1314

RESULT 14
US-09-954-456-1921
; Sequence 1921, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1921
; LENGTH: 7869
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1921

```

Alignment Scores:	
Pred. No.:	6.83e-06
Score:	176.50
Percent Similarity:	33.63%
Best Local Similarity:	20.15%
Query Match:	5.90%
DS:	9
Length:	7859
Matches:	136
Conservative:	91
Mismatches:	225
Indels:	224
Gaps:	31

```

US-09-787-016A-3 (1-562) x US-09-954-456-1921 (1-7869)

Qy      11  GluAlaProLysAlaIleLysProThrSerLysGluPheArg-IysThrItpGlyPheAr 30
      Db      ::::
      3845  GATGGCCCAAGGCGAAGAGCGCGCAACCAATCCACTGGAAGAAAGACCTGGTCGAAAC 3904
Qy      30  GArThrThrIleAlaLysArgGluGlyAlaGly-AspAlaGluAlaAsProLeuGluP 50
      Db      ::::
      3905  CAGCAITTAAGTTGAGTCGGGAAATCATCGCCACTTTCTACTCAAGCATGCGTCATTGAGC 3964
Qy      50  roProProGInGInLeuGInLysSerLeuArgArgSerGlyArgGInProLysA 70
      Db      ::::
      3965  CCATC-----GTTTCATTCTTAAGCTGGAGCTAAACCCAGA 4003
Qy      70  rgThrGluArgValGInGInPheLeuThrIleAlaArgArgGlyArgArgSerMetp 90
      Db      ::::
      4004  TCCAGAGAGCTGAAGAA-----ACTGTT-----GAGCCAAAGACAGACATGC 4045
      Db      ::::

```


FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:
 Pred. No.: 0.0343 Length: 9025608
 Score: 176.00 Matches: 140
 Percent Similarity: 34.47% Conserved: 72
 Best Local Similarity: 22.76% Mismatches: 218
 Query Match: 5.89% Indels: 189
 DB: 15 Gaps: 24

US-09-787-016A-3 (1-562) x US-10-156-761-1 (1-9025608)

```

QY 53 ProGluGlnLeuGlyLeuSerLeuArgSerGlyArgGlnProLysArgThrGlu 72
DB 5319462 CTTGCGAGCTATCAGGCGCCGAGGACCGTGCAGAGCTGATCAGCTCGGCTGGC 5319403

QY 73 Arg-----ValGluGlnPheLeuThrIleAlaArgArgGlyArgArgSerMetPro 90
DB 5319402 CGGCGGTGGCCCGACACATCATCGTCTGCGCGAGTCTCTCGACA-TGATGCGAG 5319344

QY 91 ValSerLeuGluAspSerGlyGluProThr----- 100
DB 5319343 GTGACACCTTCGCGTGCAGCTCTCGACACCCCGGTCTGTTGGGGCGGTGCGCGAGA 5319284

QY 100 ----- 100
DB 5319283 ACCGABACCACTGCACGACACCGGCCATGCAGACGCCCGCGCGCGCTCGCACC 5319224

QY 101 -----SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSer 117
DB 5319223 TGGGGCGGCCAGCTGCTCCGCGGACGAGGCGGTGGCGGAGTTCGGCCCGCGGCTCGC 5319164

QY 118 AlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValGlyGluArgProAla 137
DB 5319163 GCGACACCACTGCTAGTCTCCGCGGCGAGCTTCGCGAGTCTACTCGACGAGACTTCG 5319104

QY 138 SerSerGluLysValGlyGlyAspAspPheAspAspThrSerAsp----- 153
DB 5319103 TCAAGCGCGCACCGGCTCAAGTGGCTGAAGAGATCTTTCTATCATCTGCTGGCCCTCG 5319044

QY 154 -----Ser-AspSerAspGlyLeuThrLeuLys 162
DB 5319043 CGGTATCGGCGCGGAGACGTACGTGGCTTACCGCTGAGCGCAGACGCGAGTA----- 5318992

QY 162 sGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgProLeuLys 182
DB 5318991 ----CTAGCTCGCGCGCAGCGGAGCAGCTGCCCTGTATCCGCGGCATCAGCCAGGACC 5318936

QY 182 sGlyIleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThrValGlu 202
DB 5318935 TGGCCTGGGTCTCCTCTCGAA-----GG 5318912

QY 202 ySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAs 222
DB 5318911 TGCAGAGGACCAACCCGAGATCGACTCACTCACTGCTGCGCGGTACCGGATCAGCAGACGAG 5318852

QY 222 pGlnGly-ValValSerGlnAlaGlyLysAspArgGluSerLysLeuGluGlyLysA 242
DB 5318851 TCAAGCGCAGCATCGCGGAGCGGCTCTCAAGACCGCGAGTCAAGAGTCAAGCAGACTCG 5318792

QY 242 laalaglnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysG 262
DB 5318791 CCGTACAGGC-GTCCGCTGTAGAA-----GGACGCGCTTCGCGCG----- 5318752

QY 262 luGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnArgPheM 282
DB 5318751 -----TGCTTC-----CGAGAGCCAGAACCAACGCCAA----- 5318725

QY 282 etIleCysCysAspArg-----CysGluGluTrpPheHisGlyAspCysValGlyIleS 300

```

```

DB 5318724 -----GACCGCGAGCGCGAGCGAGGTGGGACCAACGCGGACCAACAGACCTCT 5318676
QY 300 exGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysT 320
DB 5318675 CTGACGTCCAGGCGACGCTCGTGGCC-----GACCCCGAGCCCGCTGAGAGTGGCAGCG 5318622
QY 320 hrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGln----- 336
DB 5318621 TGTCTCTCAAGTCCAAAGACCGACCTCCACACCGCGGCCCGCCAGCTCTTCGAGGAA 5318562
QY 337 -----GluAlaLysTrpArgProGlyAspAlaAspG 347
DB 5318561 GACGAGAAGCTGCTCTGCTGCTGCGGTAGACGTAAGAGCCGCTGAGGG----- 5318512
QY 347 lyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleL 367
DB 5318511 -----CCCTGTACACCATGAGCAGTACGAAACGCGCGGACGACACGCTGCACCA 5318457
QY 367 ySgIlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLys 387
DB 5318456 TGGGCTCGATCGGCGCACCGAGCGCGCACACCGAGTTGGCGTCTGCTGGTGTTCGCG 5318397
QY 387 roValIle-GluAla-----ProGlyAlaSerLysCysIle 398
DB 5318396 TGTCTATCCCGGTCTTCGGGTACGCCAACGTCGCGCTGCGCTGCTCGACGAAATGC----- 5318340
QY 399 GlyProGlyCys-----CysHisAlaLaglnProAspSer 410
DB 5318339 CGCGCGGCTCTGAGCTACGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318280
QY 411 ValTyrCysSerAsnAspCysIleLeuLysHis-AlaAlaAlaThrMetLysPheLeuSe 430
DB 5318279 GCGTACGCAAGTTGGCACCGTACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318234
QY 430 rSgIlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLys-- 449
DB 5318233 -GGCGACGCTGCTCAAGCGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318175
QY 450 ----Pro-SerLeuProLysCysGlyAlaGluAlaGlyLysLysLysLysLysLysLysL 468
DB 5318174 TGTCTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318119
QY 468 ys-----ArgProA 471
DB 5318118 GCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318059
QY 471 laProGluLysLysGluThrThrValLysLysAlaValValValProAlaArgSerGluA 491
DB 5318058 CAGCGCTACACCTACATCTCCATGTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5318008
QY 491 laLeuGlyLysGluAlaAlaCysGluSerSerThr-----P 503
DB 5318007 CCGCTGGTTC-CGGGCTCGGCGCCACACATCTACGCGCGCAGATCTGGATCTCGATTCC 5317949
QY 503 roSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAlaAlaProS 523
DB 5317948 CGGCTGGGACCTCC-----AGCCGCGGAGTTCGGGAGATCATCTCGCGGT 5317898
QY 523 erProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAspProSerA 543
DB 5317897 CTCTTCG-----CGGCTA 5317883
QY 543 rGserPheTrpIleAlaIleProTrpAlaCysPro 554
DB 5317882 CTTCTGCTCAGCGCGAGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 5317848

```

Search completed: May 1, 2004, 21:43:26
 Job time : 5588 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model
Run on: May 1, 2004, 17:18:53 ; Search time 124 Seconds
(without alignments)
2515.181 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPDSNEAPKAIKPTS.....RSFWIAIPWACPGVGVAALC 562

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DSV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09787016/runat 30042004 083703 4995/app_query.fasta_1.711
-DB=Issued Patents NA -OFMT=fastap -SUPPLY=p2n.rni -MINMATCH=0.1 -LCOPCL=0
-LCOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1.85 @runat 30042004 083703 4995 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PGTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				§							
Result No.	Score	Query Match	Length	DB ID	Description						
1	232	7.8	6256	4	US-09-620-312D-269	Sequence 269, App					
2	181	6.1	1404	4	US-09-857-556A-31	Sequence 31, Appl					
3	174.5	5.8	8931	3	US-09-051-019-1	Sequence 1, Appl					
4	170.5	5.7	4056	4	US-10-164-595-55	Sequence 55, Appl					
5	167	5.6	3228	4	US-09-252-991A-1816	Sequence 1816, Ap					
6	167	5.6	3546	4	US-09-252-991A-2143	Sequence 2143, Ap					
7	163.5	5.5	2433	4	US-09-489-039A-2577	Sequence 2577, Ap					
8	161.5	5.4	4248	4	US-10-164-595-53	Sequence 53, Appl					
9	161.5	5.4	6605	1	US-08-769-309A-4	Sequence 4, Appl					
10	161.5	5.4	6605	4	US-08-994-570-4	Sequence 4, Appl					
11	161.5	5.4	6608	4	US-09-220-132-58	Sequence 58, Appl					
12	161.5	5.4	6755	3	US-08-931-999-4	Sequence 4, Appl					

13	161	5.4	1032	4	US-09-257-179-21	Sequence 21, Appl
14	160	5.4	2344	3	US-08-893-852A-2	Sequence 2, Appl
15	158.5	5.3	13842	3	US-09-105-537-30	Sequence 30, Appl
16	158.5	5.3	36778	3	US-09-105-537-5	Sequence 5, Appl
17	158.5	5.3	38506	3	US-09-320-878-19	Sequence 19, Appl
18	158.5	5.3	38506	4	US-09-141-908-1	Sequence 1, Appl
19	158.5	5.3	38506	4	US-09-657-440-19	Sequence 19, Appl
20	157	5.3	3240	4	US-09-489-039A-331	Sequence 331, App
21	157	5.3	5173	4	US-08-801-308-2	Sequence 2, Appl
22	154	5.2	1887	4	US-09-252-991A-15843	Sequence 15843, A
23	154	5.2	2103	4	US-09-252-991A-15876	Sequence 15876, A
24	154	5.2	5200	4	US-08-978-277A-3	Sequence 3, Appl
25	152.5	5.1	4403765	3	US-09-252-991A-14550	Sequence 14550, A
26	152.5	5.1	4403765	3	US-09-103-840A-2	Sequence 2, Appl
27	152.5	5.1	4411529	3	US-09-103-840A-1	Sequence 1, Appl
28	152	5.1	4048	4	US-09-976-594-1053	Sequence 1053, Ap
29	150.5	5.0	11236	1	US-07-853-913-1	Sequence 1, Appl
30	150	5.0	9439	3	US-09-058-489-89	Sequence 89, Appl
31	150	5.0	9551	1	US-08-056-200-93	Sequence 93, Appl
32	150	5.0	9551	2	US-08-600-644-93	Sequence 93, Appl
33	149.5	5.0	2050	4	US-09-620-312D-761	Sequence 761, App
34	149.5	5.0	8146	4	US-09-976-594-725	Sequence 725, App
35	149.5	5.0	11917	4	US-09-566-921-32	Sequence 32, Appl
36	149	5.0	8438	1	US-07-945-283-1	Sequence 1, Appl
37	149	5.0	12508	4	US-09-655-270A-1	Sequence 1, Appl
38	149	5.0	12523	4	US-09-651-941-1	Sequence 1, Appl
39	149	5.0	12523	4	US-09-955-597-1	Sequence 1, Appl
40	148	5.0	2302	4	US-09-620-312D-915	Sequence 915, App
41	147	4.9	2262	4	US-09-252-991A-1060	Sequence 1060, Ap
42	147	4.9	4257	2	US-08-690-473-1	Sequence 1, Appl
43	147	4.9	4257	3	US-09-259-821A-1	Sequence 1, Appl
44	147	4.9	4257	3	US-08-843-659-1	Sequence 1, Appl
45	147	4.9	5121	4	US-09-252-991A-15189	Sequence 15189, A

ALIGNMENTS

RESULT 1
US-09-620-312D-269
; Sequence 269, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT FILING DATE: 2000-07-19
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 269
; LENGTH: 6256
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257)..(5443)
US-09-620-312D-269

Alignment Scores:
Pred. No.: 4,96e-10 Length: 6256
Score: 232.00 Matches: 94
Percent Similarity: 36.43% Conservative: 63
Best Local Similarity: 21.81% Mismatches: 154
Query Match: 7.76% Indels: 120
DB: 4 Gaps: 17

US-09-787-016A-3 (1-562) x US-09-620-312D-269 (1-6256)
Qy 92 SerLeuGluAsp----SerGluGluProThrSerCysProAlaThrAspAlaGluThrAla 110
Db 1106 TCTTTGAGTGAAGTCAACGCTCATCCTGCTGCTTGAAGAACCTCATCATCTCTGCA 1165
Qy 111 SerGluGlySerValGluSerAlaSerGluThrArg----SerGlyProGlnSerAlaSer 129
Db 1166 CAACCTGACATGATATCATCTTACGACGAAACAGTCTCATAGCCTCAGCACAGGCC 1225
Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLys----GlyGlyAspHis 148
Db 1226 CCAGCATATGAACCAATAGTCACTGAGGAGAGAGCTTGAACACCCAGGCGTTGAGCAT 1285
Qy 149 AspAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGlu-----Leu 164
Db 1286 -----TTTAAAGAGAGAGTAACTGAACCTGAAACCTGAGAGAGACCTA 1333
Qy 165 GlnAsnArgLeuArgArgGluGluGlnGluProThrGluArgProLeuLysGlyIle 184
Db 1334 CAACCCGCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1393
Qy 185 GlnSerArgLeuArgGlyLysArgArgGluGluGluGluProAlaGluThrValGlySerGlu 204
Db 1394 CCAGATAACATGATACCATCAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429
Qy 205 AlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGly 224
Db 1429 ----- 1429
Qy 225 ValValSerGlnAlaGlyLysAspArgGluSerLysLeuGluGlyLysAlaAlaGln 244
Db 1430 -----GATCATAGCTCTCTCATTTGAAAGCAATAT----- 1459
Qy 245 AspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyr 264
Db 1460 -----ATGTGACTCCAGCAGCAGCTGT---GGGTTT 1489
Qy 265 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 284
Db 1490 -----TGCAGAAACACATGTCGTGTGAAATGTTGT-----GCT 1525
Qy 285 CysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAlaArgGly 304
Db 1526 TGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1585
Qy 305 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 324
Db 1586 CAGCAGATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1636
Qy 325 GlnAspGluThrHisSerGluThrAlaAsp-----GlnGlnGluAla 338
Db 1637 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1696
Qy 339 LysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGln 358
Db 1697 GAATTCATAGTGGAGATGAACAC---ATGGAGTGTGAAGAGCTGATGATTAACAAACAC 1753
Qy 359 LysSerSerGlu-----AspGlnGlyIleLysGlyArgIleGluLysAla 373
```

RESULT 2

```

US-09-857-556A-31
; Sequence 31, Application US/09857556A
; Patent No. 6558915
; GENERAL INFORMATION:
; APPLICANT: Rebecca E. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng
; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
; CURRENT APPLICATION NUMBER: US/09/857,556A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-857-556A-31
```

```

Alignment Scores:
Pred. No.: 1,57e-06 Length: 1404
Score: 181.00 Matches: 127
Percent Similarity: 33.68% Conservative: 65
Best Local Similarity: 22.28% Mismatches: 244
Query Match: 6.06% Indels: 138
DB: 4 Gaps: 20
```

US-09-787-016A-3 (1-562) x US-09-857-556A-31 (1-1404)

```

Qy 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 14 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 73
Qy 41 GlyAspAlaGluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSer 60
Db 74 GCGTGTCTCTCAGAGACACCAATTTCTCTCCGCGCTCGAGCTCCAGGCCAAGCGCTCA 133
```


LOCATION: (189)...(2528)
OTHER INFORMATION:
US-10-164-595-55

Alignment Scores:

Pred. No.: 6,48e-05 Length: 4056
Score: 170.50 Matches: 132
Percent Similarity: 35.68% Conservative: 76
Best Local Similarity: 22.64% Mismatches: 235
Query Match: 5.70% Indels: 141
DB: 4 Gaps: 24

US-09-787-016A-3 (1-562) x US-10-164-595-55 (1-4056)

QY 7 ProSerAsnGluGluAlaProLysAlaIleLeuProThrSerLysGluPheArgLysThr 26
DB 861 CCCTCAGAGATCTGATTCCTACCCAGGAG-----AAGAAGAGAGAAAGAGGAG 914
QY 27 TrpGlyPheArgThrThrIleAlaLysArgGluGlyAlaGlyAspAlaGluAlaAsp 46
DB 915 GACATTTTCGAGATTTCAGTGGCC-----941
QY 47 ProLeuGluProProProGlnGlnGlnLeuGlyLeuSerLeuArgSerGlyArg 66
DB 942 CCACTGATCCCTTATCCACTCATCAAGGA-GGATATAATGCTATAGAAATGGAAGA 1000
QY 67 GlnProLysArgThrGlu-----ArgValGluGlnPheLeuThrIleAlaArgArg 84
DB 1001 AGACAAAGAGAGCTGATATCTCGAGGATCAACAAATTCAGAGACACATAGAAGACT 1060
QY 85 GlyArgArgSerMetProValSerLeuGluAspSerGly-----GluProThrSerCys 102
DB 1061 GGAAGAGAGAGGCGAAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 103 ProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArg 122
DB 1118 -----AGAAAGAGAGAGGAGCGGTGAAAGGGAACGAGAAAGCGGAGACGAGA 1171
QY 123 SerGlyProGlnSerAla-----SerThr 130
DB 1172 AAGGGAAGAGAGCGGGAAG 1231
QY 131 Ala-ValLysGluArgProAlaSerSerGluLysValLysGlyGlyAspHisAspAs 150
DB 1232 GCAGAGAGAGAGCGGAG 1291
QY 150 pThrSerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArg 170
DB 1292 CCAGAGAGAGAGATGAAGAGAGATGCTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
QY 170 GlyArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgL 190
DB 1352 GAAA-----GAAGCTGTTATCAGAGAGCG-----CTTAAGAAATGGGAAATCAGAGAGAGAG 1405
QY 190 slys-----ArgArgGluGluGlyProAlaGluThrValG 202
DB 1406 GAAACCCGGGATATGAG 1465
QY 202 ySerGluAla-----SerAspThrValGluGlyValLeuProSerLysGlnGluPr 219
DB 1466 CAAGAGAGCTAAACGACTAAAGAAATCTTAGAGAGACTATGATGATGATGATGATGAT 1525
QY 219 oGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuG 239
DB 1526 CAAATATTACAGAGAGAGTGTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1585
QY 239 uGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPr 259
DB 1586 AGCA-----GATGAG 1618
QY 259 oGluCysGluGlyTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAs 279
DB 1619 CGAGCTTGAG-----1628

QY 279 nargPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyTl 299
DB 1628 -----1628
QY 299 eSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCy 319
DB 1629 -----GAATCAGGAGCGCTTCTGCGAGAGGCGATCCAGAT-----CCAGATCC 1675
QY 319 sThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAla 339
DB 1676 AGAGCTCCAGAGGATGAGCAAGAG-----GCTCAGAGGCGCAGGAGCAGCAATAAA 1729
QY 339 sTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGln 359
DB 1730 GCAGAGCGCGAATCAGAGAGAGAGAGAG-----GAAGAGAGAGAGAGAGAGAGAG 1765
QY 359 sSerSerGlnGlnGlyIleGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
DB 1766 GCAG 1825
QY 379 sLysLysLysLys-----IlePheGlnProValIleGluAlaProGlyAlaSerLys 398
DB 1826 GCCTTGTCTGAAACCTACTCTGAGCGCCATCAGCTCTGCTCCATCTGTCTCTGCTCC 1885
QY 398 eGlyProGlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCys 417
DB 1886 TGCGAAT-----GCAACCTAACCTCTCTGGGATGAGTCTCCCTGTGG 1930
QY 418 -----IleLeuLysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGln 436
DB 1931 TATTATTATCTCATGAA-----AACTCACAGATCAACAGCA 1969
QY 436 sPro-----LysProLysGluLysMetLysMetLys-----ProGluLysPr 450
DB 1970 ACCTGAGGAGCATAGCGCAAAATAGGACTAAGCTTAACTTGGGTGCTTCCATAGTCC 2029
QY 450 oSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisArgPr 470
DB 2030 TGCTCAGCTTAATCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2089
QY 470 oAlaProGluLysLysGluThrThrValLysLysAlaValValValProAlaArgSer 490
DB 2090 TGAGATGATGAGACAGTGTAGTACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2149
QY 490 uAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHis 510
DB 2150 TGAAGATGATAAATGCAACCAAGGCACTGTAAACACTGAAGAGAGAGAGAGAGAGAGAG 2209
QY 510 nTyAsnAlaValLysProGluLysThrAlaAlaPro-SerProSerLeuLeuTyLysC 530
DB 2210 TAAGAGTCTCTAG 2262
QY 530 yMetTyHisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaLleP 550
DB 2263 -----CCCTGGATGTGCTATTG-----TGGATCTATACT 2293
QY 550 rofTrp 551
DB 2294 GATGG 2298

RESULT 5

US-09-252-991A-1816
Sequence 1816, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 1816
;; LENGTH: 3228
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1816

Alignment Scores:

Score: 9,336-05 Length: 3228
Pred. No.: 167,00 Matches: 140
Percent Similarity: 36.93% Conservativeness: 86
Best Local Similarity: 22.88% Mismatches: 217
Query Match: 5.59% Indels: 172
DB: 4 Gaps: 30

US-09-787-016A-3 (1-562) x US-09-252-991A-1816 (1-3228)

QY 4 LysGlyAspProSerAsnGluAlaPro----LysAlaIleLysProThrSerLysGlu 22
DB 1553 AAATGGCCACCGTCGAGCAGGAAGCCACCGCGTCTGAGTCTGAGCCCGACCCCTGGTCC 1612
QY 23 PheArgLysThrTyrGly-PheArgArgThrThrIleAlaLys----- 36
DB 1613 GCCAGAGAGCAGCAGTCAAGACCGTCTGCTCCCGACGAGCCCGCACCAACACCGAAG 1672
QY 37 -----ArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProProProGlu 54
DB 1673 CACCGGTCTGAGCCGCGCAGCGCGATGCCCGAGCCGAGCTGT-----CCAGGCCCTGG 1726
QY 54 nGlnGlnLeuGlyLeuSerLeuArgArgSerGly-----ArgGln-ProLysArgT 71
DB 1727 TGAAGTCCCTGCTGGCTGCTGTTGCGAGGCAAGATCACTGCGCCAGCTCTGTGAAA 1786
QY 71 hrGluArg-----ValGluGlnPheLeuThrIleAlaArgArgGlyArgArgS 88
DB 1787 CCAGCAAGCCGCTGCTCGAGCGCCAAACCCCGCAGGAGCGGTGCGCAACCGCCCGCAGC 1846
QY 88 erMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaG 108
DB 1847 AG-----AACCGCCCGCGATGCGCGATGG-CAATCGCGCGCGACGAGAGCGC 1896
QY 108 luThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyPro----- 125
DB 1897 AAACCGCG--CGAGGAGC-GTGACAGCGTCAACCGCGGAGAGCGCGCGACCGCC 1952
QY 126 -----GlnSerAlaSerThrAlaValLysGluArgProAlaSerS 139
DB 1953 GAACCGCGAAGCGCGAGCGCGCGCGCGGAGAGCGCGCC-GAGCGTCCGCTCGCG 2011
QY 139 erGluLysValLysGlyLysAspHisAspThrSerAspThrSerAspSerAspGlyL 159
DB 2012 AGGAGCGC----- 2019
QY 159 euThrLeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluA 179
DB 2020 -----CAGCGCGCGAAGCGCGTGAAGAGCGCGCGCGAGCACCGCC 2062
QY 179 rg-----ProLeuLysGlyIleGlnSerArg--LeuArgLysLysArgArg 194
DB 2063 GCGAAGAGCGCCACCGCGGAGCGCGGAGGTGCGGAGGACCGACCGAAGCGCGCC 2122
QY 194 luGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuP 214
DB 2123 GCGAAGAGCGCGCGAAGCGCGCGCGCGCGAAGAGCGCGCAACCGCGAAGCGCGTGAAG 2182
QY 214 roSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspA 234
DB 2183 AACCGTCTGAGACCGCCCGCGCGCGAGAG-----CGCCAGCGCGCGGAGATC 2230
QY 234 rgGluSerLys-----LeuGluGlyLysAla-----A 243

RESULT 6
US-09-252-991A-2143/c

DB 2231 GCCAGGCTCGAGACGCGCGCGCCCTGGAGCCCGAGGCACTTCCGCAACGACGAGAGCTGG 2290
QY 243 laGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluG 263
DB 2291 AGCAGGACGACGACGAGGATACCGATGCGGAG-----CGCCCGCGCGCGCGCTCCGCG 2344
QY 263 lyTyrrAspProAsnAlaLeuTy-CysIleCysArgGlnProHisAsnAsnArgPheMetI 283
DB 2345 GCCAG-----CGTGTCTCGACGACGACCGCGCG----- 2370
QY 283 leCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSerGluAlaA 303
DB 2371 -----GAACGCGACGCGCGAG---GTACGCGCGAGGAGGTGGAAGGC---AGCAGGCGGA 2416
QY 303 rgGlyArgLeuLeuGluArgAsnGlyCyluAspTyrrIleCysProAsnCysThrIle--- 321
DB 2417 CC-----GATATGCGCGCGCGCGCTGCAACACCTGTCCGACG 2452
QY 322 -----LeuGlnValGlnAspGlu----- 327
DB 2453 CCGCGCGCGTCTCGCGTATCTGTCTCCAGAGGCGGTAGAACAAACGTGAGCAAG 2512
QY 328 -----ThrHisSerGluThrAlaAspGlnGlnGluAlaLysTyrPargProGlyAspA 345
DB 2513 CCGCGCGCCTTACCGAGCGAGGCTCCAGCGAAACCAACCGCA-----A 2554
QY 345 laAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnG 365
DB 2555 GCGATGAGACGACGCGCTCGACCGAGCGCGTGTGCGAAGCCAGCGCGACGAGCGAG 2614
QY 365 lyIleLysGlyArg-----IleGluLysAlaAlaAsnP 376
DB 2615 CCAATACCGCGGAAACCGCGACATCGAAGCGCGCGGTGACCGTCTGCGTGTCTCGGAGCG 2674
QY 376 roSerGlyLysLysLysLysIlePheGlnProValIleGluAlaProGlyAlaSerL 396
DB 2675 AAGCGACGACGACGACCTGTGTGCGCGAAGCCACCTGAGAGAGCTCCCTTCGCCACG 2734
QY 396 yscYsIle----- 398
DB 2735 AAAGCGTGAAGACCGCGAGACGCGGAGCGCGCTGCAACCGCGAAGCGCGCG 2794
QY 399 -----GlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerA 415
DB 2795 AAGAGTGTGCGTCTCGCGTGTGCGAGTAGCAGCCCTTAGCGAGCGCGCGACCG 2854
QY 415 snAspCysIleLeuLysHisAlaAla-----AlaThrMetLysPheLeuSerS 431
DB 2855 AGGAGCGACCGCGCGCATCGCGGCGTCCGCGCGCAACCGACCTGCGCGTCTCAACG 2914
QY 431 erGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProS 451
DB 2915 ACCCAGCGGAAAGCGTCTGCGACGCGAAGCGCGGCTGTGCGACGCGAGCGCGAG 2974
QY 451 erLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProA 471
DB 2975 CAGCAGCGGAGCGCGACCTCAGCGCGT-----CCCGCGTCTGAGGAGATCCCGG 3025
QY 471 laProGluLysGluThrThrValLysLysAlaValValProAlaArgSerGluA 491
DB 3026 CTGTAGCGGAGGAGGAGCGTCTGCGCGCGAGGAGAA-----CCTGTGCGACCGCCAGG 3076
QY 491 laLeuGlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerAspHisAsnT 511
DB 3077 CT-----GAGAAATACCGACGCGCGAGCTTCTGCGCGCGCGCGCGCGCG 3124
QY 511 yrAsnAlaValLysProGluLysThrAlaAla 521
DB 3125 AGGAAGCGGTACGCGCGCGAGCTTGAAGCTTCC 3156


```
; Sequence 2143, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2143
; LENGTH: 3546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2143

Alignment Scores:
Pred. No.: 0.000107 Length: 3546
Score: 167.00 Matches: 140
Percent Similarity: 36.93% Conservative: 86
Best Local Similarity: 22.88% Mismatches: 217
Query Match: 5.59% Indels: 172
DB: 4 Gaps: 30

US-09-787-016A-3 (1-562) x US-09-252-991A-2143 (1-3546)
Qy 4 LysGlyAspProSerAsnGluGluAlaPro---LysAlaIleLysProThrSerLysGlu 22
Db 1814 AAATGGCCACCGTGCAGCAGCAAGAACGCCAGCGCGTGCAGTCGACCCGCCAOCCTGGTCC 1755
Qy 23 PheArgLysThrTrpGly-PheArgLysThrThrIleAlaLys----- 36
Db 1754 GCCAGGAGCAGCAGTCAAGACCTGCTCCCGAGCGCCGACCGCACACACCGAAG 1695
Qy 37 -----ArgGluGlyAlaGlyAspAlaGluAlaAspProLysGluProProGlu 54
Db 1694 CACCGTGCAGCGCCAGCGCATGCGCGAGCGCGCGTCTT-----CCAGGGCCCTGG 1641
Qy 54 nGlnGlnLeuGlyLeuSerLeuArgArgSerGly-----ArgGln-ProLysArgT 71
Db 1640 TGAAGTCCCTGCTGCGCTGTTCGAGCAGCAGGATCACTGCGCCGCAAGCTGCTGAAA 1581
Qy 71 hrGluArg-----ValGluGlnPheLeuThrIleAlaArgArgGlyArgArg 88
Db 1580 CCACCAAGCGCGCTGCGAGCGCCAAACCCCGAGCAGCGTCCGCAACCGCGCGCAGC 1521
Qy 88 erMetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaG 108
Db 1520 AG-----AACCGCGCGCGATGCGCGCGATGG-CAATCGCGCGAGCAAGAGCGC 1471
Qy 108 luThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyPro----- 125
Db 1470 AAAACCGCG---CGAGGAGC-GTGAGAGGTCACCGCGCGAGAGCGCGCGCAACGCC 1415
Qy 126 -----GlnSerAlaSerThrAlaValLysGluArgProAlaSerS 139
Db 1414 GAACCGCGAAGAGCGACGCAACCGCGCGCGAGAGCGCGCC-GAGGTCGCGCTCGCG 1356
Qy 139 erGluLysValLysGlyAspAspHisAspAspThrSerAspSerAspGlyL 159
Db 1355 AGGAGCGC----- 1348
Qy 159 euThrLeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluA 179
Db 1347 -----CAGCGCGCGAGAGCGCGTGAAGAGCGCGCGCGCGAGCGCGCACCCC 1305
Qy 179 rg-----ProLeuLysGlyIleGlnSerArg---LeuArgLysLysArgArg 194
Db 1304 GCGAAGAGCGCAACCGCGCGAGCGCGCGAGGTCGCGAGGACGAGACGCGAGCGCGCGCC 1245
```

```
Qy 194 luGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuP 214
Db 1244 GCGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
Qy 214 roSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspA 234
Db 1184 AACCGGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 1137
Qy 234 rgGluSerLys-----LeuGluGlyLysAla-----A 243
Db 1136 GCGAGCTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1077
Qy 243 laGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluG 263
Db 1076 AGCAGGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1023
Qy 263 lyTyraAspProAsnAlaLeuTyrcysileCysArgGlnProHisAsnAsnArgPheMetI 283
Db 1022 GCCAG-----CGTGTGCGAGCAACCGCGC----- 997
Qy 283 leCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaA 303
Db 996 -----GACGCGCAGCGCGAG---GTCAGCGCGCGAGGTGGAAGGC---AGCGAGCGCA 951
Qy 303 rgGlyArgLeuLeuGluArgAsnGlyGluAspTyrlleCysProAsnCysThrIle----- 321
Db 950 CC-----GATAATGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 915
Qy 322 -----LeuGlnValGlnAspGlu----- 327
Db 914 CCGCGCGCGCTGCGGTATGCTGTGCGAGCGAGCGCGTAGAAGCAACAGTGGAGCAAG 855
Qy 328 -----ThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArgProGlyAspA 345
Db 854 CCGCGCGCCTACCGAGCGAGGTGCGAGCGAACCACCGCA-----A 813
Qy 345 laAspGlyThrAspCysThrSerIleGlyThrIleGlnLysSerSerGluAspGlnG 365
Db 812 GCGATGAGCGCGCGCGTGCAGCGAGCGCGTGCAGAACCGCGAGCGCGCGAGCGCGAGG 753
Qy 365 lyIleLysGlyArg-----IleGluLysAlaAlaAsp 376
Db 752 CCAATACCGCGCAACCGCGACATCGAGCGCGGTGACCGTCAGCGTGTTCGGGAGCG 693
Qy 376 roSerGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerL 396
Db 692 AACCGGACGAGACCGCTGCTGTCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
Qy 396 ysCysIle----- 398
Db 632 AAAGCGTGAAGCGCGAGAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
Qy 399 -----GlyProGlyCysCysHisValAlaGlnProAspSerValTyrcysSerA 415
Db 572 AAGAAGTTCGCTGCGGTGCGCGTGCAGTAGCAGCGCCCTAGCGAGCGCGCGCGCGCGCG 513
Qy 415 snAspCysIleLeuLysHisAlaAla-----AlaThrMetLysPheLysSerS 431
Db 512 AGAGCGCGACCGCGCGCATCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
Qy 431 erGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProS 451
Db 452 ACCCAGCGGAAACCGCGCGTGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
Qy 451 erLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProA 471
Db 392 CAGCGCGGAGCGCGCGAGCTCAGCGCGCT-----CCGCGCGTGCAGAGATCCCGG 342
Qy 471 laProGluLysLysGluThrThrValLysLysAlaValValProAlaArgSerGluA 491
Db 341 CTGTAGCGCGCGAGGAAGCGTCCGCCCGCAGGAGAA-----CCTGCTGCACCCCGG 291
```

Qy	491	lateuGlyIysgluAlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnT	511
Db	290	CT-----GAAGAAATCACCCAGGCGCGAGTTCGTCCTCCAGGCG---	GACGAAGGCC 243
Qy	511	yrAsnAlaValIysProGluIysThrAlaAla	521
Db	242	AGGAAGCGGTACAGGCGCGAGCTCGAAGCTTCC	211

QY 520 -----AlaIaIaProSerProSerLeuLeuTyriLysCysMetTyriHisL 534
Db 384 CACGCTCCGCGCTGGCGCTGCACCCAGCCATCATC 337
QY 534 euGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysP 554
Db 336 CTGGTAGCGCTTC-----CGCGTAAACACGAGATGTTCCAGCAGGTGGCC 289
QY 554 roGlyLeuGlyValAlaAlaLeu 561
Db 288 CAGCCCGAGGAGCGCGAGCGTT 266
RESULT 8
US-10-164-595-53
; Sequence 53, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-04-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 4248
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (189)..(2720)
; OTHER INFORMATION:
US-10-164-595-53
Alignment Scores:
Pred. No.: 0.000428 Length: 4248
Score: 161.50 Matches: 129
Percent Similarity: 34.06% Conservative: 75
Best Local Similarity: 21.54% Mismatches: 246
Query Match: 5.40% Indels: 150
Gaps: 24
US-09-787-016A-3 (1-562) x US-10-164-595-53 (1-4248)
QY 2 AspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSerLys 21
Db 999 GAAGACAAAGAGACCTGATATCTCGAGAGATCAGCAAA----- 1037
QY 22 GluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAlaGly 41
Db 1038 ---TTCAGAGACACA-----CATAGAAACTGGAGAGAGAGAAAGGCAAAAG 1082
QY 42 AspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuSerLeu 61
Db 1083 GAAAGAAAGACAGCAAGAAATTGAGAAAGACGAGAGAGAGAGAGAGGAGCGTGAAAGG 1142
QY 62 ArgArgSerGlyArgGlnProLysArgThrGluArg-ValGluGlnPheLeuThrIleAl 81
Db 1143 GAACGAGAAAGCGCGAGAGCGGGAACGA---GAAAGGGAAGAGAGACGTGAACGAGAAAG 1199
QY 81 aArgArgArgGly----- 85
Db 1200 GAGAAAGACGGGCGGACGAGAAACGGGATAGGACCGTGACCGGACAAAGAGAGA 1259
QY 86 -----ArgArgSerMetProValSerLeuGluAspSerGlyGluProTh 100
Db 1260 GACCGAGATCGGATCGAGAGAGAGATCGTGACCGGATAGAGAAAGGAGCTCAGATCGT 1319
QY 100 rSerCysProAlaThrAspAlaGlu-----ThrAlaSerGluGlySe 114
Db 1320 AATAGGATCGCGTCGATCAGAGAGAAAGAAAGCAGAGATCGTGAAAGGAGGAGAGCGG 1379
QY 114 rValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValLysG 134

Db 1380 GAAAGAGAGAGAGAGAGACGAGACGAGAACGAGAACGAGAACGAGAACGAGAGAGAGAG 1439
QY 134 uArgProAlaSerSerGluLysValLysGlyGlyAspAspHisAspAspThrSerAspSe 154
Db 1440 GAGAGGAAACGGGAGCGAGAGAGAGAAAA---GACAAAAAAGCGGACCGAGAGAGAGA 1495
QY 154 rAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgLysArgGluG 174
Db 1496 TGAAGAGATCATACGAAACGAGAAACTTGAAGAAACTCCGAGAGAAA---GAAAGC 1552
QY 174 nGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLys 191
Db 1553 TGCTTATCAGAGCGC---CTTAAAGATTGGGAATCAGAGAACGAGAAAGAAACCCGGGA 1609
QY 192 -----ArgArgGluGluGlyProAlaGluThrValGlySerGluAla-- 205
Db 1610 ATATGAGAAAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTAA 1669
QY 206 -----SerAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspG 223
Db 1670 ACGACTAAAGAAATCTTAGAAGACTATGATGATGATGATGATGATGATGATGATGATGAT 1729
QY 223 nGlyValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAl 243
Db 1730 AGGAAGTCTCTTCAGAAAAGGTTGCGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1781
QY 243 aGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluG 263
Db 1782 -----GATGAACGA-----GATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1820
QY 263 yTyrAspProAsnAlaLeuTyriCysIleCysArgGlnProHisAsnAsnArgPheMetI 283
Db 1820 ----- 1820
QY 283 eCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyLysSerGluAlaAr 303
Db 1821 -----GAAATCAG 1828
QY 303 gGlyArgLeuLeuGluArgAsnGlyGluAspTyriIleCysProAsnCysThrIleLeuG 323
Db 1829 GCAGCGCTTCTGGCAGAGAGGCGATCCAGAT-----CCAGTCAGAGAGTCCAGAG 1879
QY 323 nValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArgProG 343
Db 1880 GATGGAACAAGAG-----GCTGAGAGCGCAGCAGCGCCACAAATAAAGCAAGAGCCAGA 1933
QY 343 yAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAs 363
Db 1934 ATCAGAGAGAGAGAGAA-----GAGAAAGCAAGAAAGAAAGA 1969
QY 363 pGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLy 383
Db 1970 AGAAAAACGAGAGAACCCATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
QY 383 s-----IlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCy 402
Db 2030 ACCTACTCTGAGGCGCATCAGCTCTGCTCCATCTGTTCTCTCTCTCTCTCTCTCTCTCTCT 2084
QY 402 sCysHisValAlaGlnProAspSerValTyriCysSerAsnAspCys-----IleLeuLy 420
Db 2085 -----GCAACACCTTAACCTCTCTGCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 2134
QY 420 sHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysPro----- 437
Db 2135 TCATGAA-----AACTCACCAGATCAACAGCAACCTGAGGAGCA 2173
QY 438 -LysProLysGluLysMetLysMetLys-----ProGluLysProSerLeuProLy 454
Db 2174 TAGGCCAAATATAGGATAGTCTTAACCTGGGTCTTCCATAGTCTCTGTCTGAGCTTAA 2233
QY 454 sCysGlyValaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLy 474

Db 2234 TTCTGTGAAGAGAAAGAACTACTGTAGATAGTGTCTTAAACAAATTTGAGGATGAAGA 2293
 Qy 474 siysgluThrVallyslveAlaValValValProAlaArgSerGluAlaLeuGlyLY 494
 Db 2294 CAGTGATGAGTACCCCAAAAGAAAGAACTGGTTCCTTGGATTTATGCTGAAGATGATAA 2353
 Qy 494 sgluAlaAlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaVa 514
 Db 2354 AAATGCAACCAAGGCACTGTAACACACTGAAGAAAGCGTAAACACATTAAGAGTCTCAT 2413
 Qy 514 llyProGluLysThrAlaAlaPro-SerProSerLeuLeuTyrLysCysMetTyrHisL 534
 Db 2414 TGAGAAATCCCTACAGCAAACTGTGCTTTCCTTCATC-----C 2455
 Qy 534 euGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrp 551
 Db 2456 CCGGATTGCTCTATTG-----TGCATTCTATCTACTGATGG 2490

RESULT 9

US-08-769-309A-4

; Sequence 4, Application US/08769309A

; Patent No. 5741890

; GENERAL INFORMATION:

; APPLICANT: Scott, John D.,

; APPLICANT: Nauert, Brian J.,

; APPLICANT: Klaucek, Theresa M.

; TITLE OF INVENTION: Protein Binding Domains of Gravin

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower/233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/769,309A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: No 5741890and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 27866/33451

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6605 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 192..5531

; US-08-769-309A-4

Alignment Scores:
 Pred. No.: 0.000832 Length: 6605
 Score: 161.50 Matches: 133
 Percent Similarity: 30.51% Conservative: 75
 Best Local Similarity: 19.42% Mismatches: 221
 Query Match: 5.40% Indels: 255
 DB: 1 Gaps: 28

US-09-787-016A-3 (1-562) x US-08-769-309A-4 (1-6605)

Qy 3 AspLysGlyAspProSerAsnGluAlaPro-----LysAlaIleLysProThrSer 20
 Db 987 GAGGAGGAGAGAGAAACCAAGAAACCTAGCAAGCTCTGCAGAACTCTCCGACTAGT 1046
 Qy 21 -----LysGluPheArgLysThrTrp-----Gly 28
 Db 1047 CCGGTGACCAAGTGAACAGGATCAACCTTCAAAAAATTTCTCACTCAAGGTGGGCGCCG 1106
 Qy 29 PheArgArgThrThr-----ProAlaGluThrValGluPro----- 33
 Db 1107 TGGCGCAAAAGACCACTTTTCAGSNAGCCGAGGAGGATGAAGTGAAGCTTTCAGAGAG 1166
 Qy 34 -----IleAla 35
 Db 1167 AAAAAGAACACAGAGCCAGAAAAAGTAGACACAGAAAGAAAGGAAAGGAGGTTGCC 1226
 Qy 36 LysArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProPro----- 51
 Db 1227 TCCGAGAACTGACCGCTCCGAGCAAGCCACCACACAGGAGCCGCGGAGAAAGTGCAC 1286
 Qy 52 -----ProGlnGlnGlnLeuGly 58
 Db 1287 GAGCCCGCTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTC--- 1343
 Qy 59 LeuSerLeuArgArgSerGlyArgGln-----ProLysArg 70
 Db 1344 -----AGTGGCTCCAGGAGCTTCTCAGAGAAACCTGCTCCGTTGGCG 1388
 Qy 71 ThrGluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgArgSerMetPro 90
 Db 1389 ACAGAAGTGTGTGATGAAAAATAGAAAGTCCACCAAGAGAGGTGTGGCCGCAAGTCCAC 1448
 Qy 91 ValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAla---GluThr 109
 Db 1449 GTCAGCACCCGTGGAG-----GAGAGAACCAAGACGAGAAACGAGGTGGAGAAACA 1502
 Qy 110 AlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 129
 Db 1503 GCAGGCTGTGCGCAGCTGAAGATTTGTTGGAATGATGATGAGAACTCTCAGAAAGCCGAA 1562
 Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyLysAspHisAsp 149
 Db 1563 CCTGCC---AAGGAGCTGTGGAAGCTCAAGAAAGCGTGTGTTCCGAGAGACCTTACA 1619
 Qy 150 AspThrSerAspSerAspGlyLeuThrLeuLysGlu----- 163
 Db 1620 CAGGAGCTGACCTCAGTCTCTGATGAGAGGTGCTGTCCAAACCCCGAGGCGTTGTG 1679
 Qy 164 -----LeuGlnAspArgLeuArgLysArgGluGlnGluPro----- 176
 Db 1680 AGTCAGGTGGAAATGCTGTCTATCACAGGAGAGAAATGAAGGTGCGAGGAGTCCACTAAG 1739
 Qy 177 -----ThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLysArgArg 193
 Db 1740 AAGCTTTTACCAGCACTGGCTTAAAGAGCTTTCTGGAAGAAACAGAAAGGGAAGAA 1799
 Qy 194 GluGluGly-----ProAlaGluThrValGlySer 203
 Db 1800 GGAGGAGGAGACGAGGAATCAGGGGAGCAGACACTCAGGTTCCAGCCGATTTCCGAGACG 1859
 Qy 204 GluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu----- 220
 Db 1860 CAGGAGGAGCAAAAGGCGAGAGCTCTGCTCATCCCTGAGGAGCCGAGGAGATCAAG 1919
 Qy 221 -----AsnAspGlnGlyValValSerGlnAlaGlyLys 231
 Db 1920 TGTCTGGAAGAGGGCTTAGCCGAGGTGAGCAGGATGGGAGAGCTGAAGAGGAGTACT 1979
 Qy 232 AspAspArgGluSerLysLeuGluGly----- 240
 Db 1960 TCCGATGAGAGAAAAAGAAAGAGAGGTGTCTACTCCCTGGGATCATTCATCAAAAGATGTG 2039

Db 1287 GAGCCCCGGTATACGTGATATAGAAAGTTGAGTGCCTCTCAGGAGCAAGTC--- 1343
Qy 59 LeuSerLeuArgArgSerGlyArgin-----ProLysArg 70
Db 1344 -----AGTGGCTCGCAGGACCTTCTGAAGAGAAACCTGCTCGGTGGCG 1388
Qy 71 ThrGluArgValGluGlnPheLeuThrIleAlaArgAArgGlyArgSerMetPro 90
Db 1389 ACAGAAAGTGTGTGATGAGAAATAGAGTCCACCAAGAAAGAGTGTGTCGCGAAGTCCAC 1448
Qy 91 ValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAla---GluThr 109
Db 1449 GTACACCGCTGGAG-----GAGAGAACCGAGAGCAGAAACGAGGTGGAGAAACA 1502
Qy 110 AlaSerGluGlySerValGlnSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 129
Db 1503 GCAGGCTGTGTCACAGCTCAAGAAATGTTGATGATGATGATGATGATGATGATGATGATGAT 1562
Qy 130 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyAspAspHisasp 149
Db 1563 CTGTC-----AAGAGCTGTGAGCTCAAGAAACGTGTGTTCCGAGAGACCTTACA 1619
Qy 150 AspThrSerAspSerAspSerAspGlyLeuThrLeuLysGlu----- 163
Db 1620 CAGGAGCTGACCTCAGTCTGATGAGAAAGTGTGTCACAAACCCCGAAGCGGTGTG 1679
Qy 164 -----LeuGlnAsnArgLeuArgArgGlyArgGluGlnGluPro----- 176
Db 1680 AGTGAGGTGGAAATGCTGTATCATCAGAGAGATGAAGTCCAGGAAGTCCACTAAG 1739
Qy 177 -----ThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysArg 193
Db 1740 AAGCTTTTACAGCAGCTGCTTAAAGCTTCTGGAAGAAACAGAAAGGAAAGA 1799
Qy 194 GluGluGly-----ProAlaGluThrValGlySer 203
Db 1800 CGAGGAGGACGAGGAATCAGGGAGCACATCAGGTTCCAGCGGATTCCTCCGACAGC 1859
Qy 204 GluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu----- 220
Db 1860 CAGGAGGACCAAAAGCGGAGAGCTGCTGCTCATCCCTGAGAGCCCGAGAGATCAG 1919
Qy 221 -----AsnAspGlnGlyValValSerGlnAlaGlyLys 231
Db 1920 TGCTCGAAAAGGGCTTAGCCGAGTGCAGAGGATGGGAGAGCTGAAGAGGAGCTACT 1979
Qy 232 AspAsnArgGlySerLysLeuGluGly----- 240
Db 1980 TCCGATGGAGAGAAAPAAAGAGAGAGTGTCACTCCCTGGGATCATTCAAAAGATGGTG 2039
Qy 240 ----- 240
Db 2040 ACGCCCAAGAGCGTGTAGACGGCTTCGAAAGTGTATAAGAGAGATGAGCTGGACAG 2099
Qy 241 -----LysAlaAlaGlnAspIleLysAspGluGlu 250
Db 2100 GTCAAGAGCGCTACCTGTGTTCCACCGAGAGCACAGCTCTGAAATGCAGAGAAATG 2159
Qy 251 ProGlyAspLeuGlyArgProLysProGluCysGluGlyTyraAspProAsnAlaLeuTy 270
Db 2160 AAAGGAGCGTGAAGAGCCAAAGCCGAA----- 2189
Qy 271 CysIleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGlu 290
Db 2190 -----GACCAAAAGCGCAAG-----GTGGATACCTCAGTATCT 2222
Qy 291 TrpPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsn 310
Db 2223 TGGAGAGCTTTAATTTGTGGATCATCCAGAAAGAGCAAGG-----AGAAGTCC 2276
Qy 311 GlyGluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSer 330
Db 2277 TCTTCTCAT-----GAGAGAGGGGACCAAAAGCA 2306

Qy 331 GluThrAlaAspGlnGlnGluAlaLysTrpArgProGlyAspAlaAsp---GlyThrAsp 349
Db 2307 ATGGGAGGAGACACCCAGAAAGCTGATGAGCCGCGAAAGACAAAGAGCGGAGCAGAC 2366
Qy 350 CysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArg 369
Db 2367 GGGATCTTCTGCTGCTCC-----CAAGAACATGATCCAGCGCAGGCA---AGTTCCTCC 2417
Qy 370 IleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProValIle 389
Db 2418 CCGGAGCAAGCTGGAGCCCTTACC----- 2441
Qy 390 GluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAsp 409
Db 2442 GAGCGGAGCGCGTTC----- 2459
Qy 410 SerValTyrCysSerAsnAspCysIleLeuLysHisAlaAlaThrMetLysPheLeu 429
Db 2460 -----ACCTGGGAGTCAATTTAAAGGTTA 2483
Qy 430 SerSerGlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLys 449
Db 2484 GTACG-----CCAGAAAAAATCAAACTCCAGCTGAGAG 2522
Qy 450 ProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArg 469
Db 2523 AAAAGCGAAGACTCCATA---GCTGGGTCTGGTGTAGACATTCCTCCACAGACTCAA 2579
Qy 470 ProAlaProGlnLysLysGluThrThrValLysLysAlaValValProAlaArgSer 489
Db 2580 CCCGTTAAAGAGAAATCTGGTCTCAATCAAGAAAGTTTATCTCTGACGAGAGAGAA 2639
Qy 490 GluAlaLeuGlyLys-----GluAlaAlaCysGluSerSerThrProSerTrpAlaSer 507
Db 2640 AGGCCAGATGGAAACAAAGAACAGCCCTCTTGAAGACGCGGCGCAACAGCGGCCAAC 2699
Qy 508 AspHisAsn-----TyrAsnAlaValLysPro 516
Db 2700 GAGATGACTCTGATGTCCCGCGTGTCTCTGTGTGATGATGATGCTGTAGAAAGG 2759
Qy 517 GluLysThrAlaAla 521
Db 2760 GAGAAATGGAGGCA 2774

RESULT 11

US-09-220-132-58
; Sequence 58, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 6608
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-58

Alignment Scores:
Pred. No.: 0.000833 Length: 6608
Score: 161.50 Matches: 133
Percent Similarity: 30.51% Conservative: 76
Best Local Similarity: 19.42% Mismatches: 221


```

QY 32 ThrThrIleAlaLysArgGluGluValAlaGlyAspAlaGluAlaAspProLeuGluProPro 51
Db 592 GCAACCAAGTCTCCTAGA---GGCAGGAGAGTCAATTTCCAGATGCCAGCGTGTCTCC 648
QY 52 ProProGlnGlnGlnSerLeuArgArgSerGlyArgGlnProLysArgThr 71
Db 649 CTGTCTCCAGCCCTCTGATAAGCACTGCAAGTCTTCGATAAGAACCCAGGGAGGAG 708
QY 72 GluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgArgSerMetProVal 91
Db 709 AAAGCCGAGAGAGGAGTGTCTGAGAGGAGGAGTTAACAGATTCTTATCCACCA 768
QY 92 SerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGluThrAlaSer 111
Db 769 TCACACCGGAG-----TGTTGTCCAGCCGTGAGGAGGAGGACCATGAA 813
QY 112 GluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAla 131
Db 814 GAAGCTGTAAAGAAAGAGCTCACAAACTCTACTTCTGCTTCTCCAGATCC--- 870
QY 132 ValLysGluArgProAlaSerSerGluLysValLysGlyGlyAspAspHisAspThr 151
Db 871 -----AAGCCCAAGCACTTGGTGTCTTGGCCAGGGGAGGAGAGATCAAGCCAG 921
QY 152 SerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgGly 171
Db 922 GAGGATAAAAGACAGAA-----AGAGTAAGAGGAGCCAGGAAG 960
QY 172 -----ArgGlu-GlnGluProThrGluArgProLeuLysGlyIleGlnSe 186
Db 961 ACCTCGGTGTCCTCCCGATCTTCAGGCTCCAGCCAGGCTCTGCGAGTATCGTTCAGGA 1020
QY 186 rArgLeuArgLysLysArgGluGluGlyProAlaGluThrValGlySerGluAlaSe 206
Db 1021 GAGG---CCTCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1077
QY 206 rAspThrValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlyValVa 226
Db 1078 GCCCCAGGCGCAATCCTCAGCCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1136
QY 226 lSerGlnAlaGlyLysAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIl 246
Db 1137 CCAACCCAGTGTGAAGAGGAGAGTGAAGGCTTTGGGGGAGGAGGAGGAGGAGGAGGAGGAG 1196
QY 246 eLysAspGluGluProGlyAspLeuGlyArgProLysProGluGlyGluGlyTyraSpPr 266
Db 1197 AGAAGCTGAGTGTCTCCCTGCAFC-----CCCCACCAAG-TGCTTT-----CC 1240
QY 266 oAsnAlaLeuTyrcyleCysArgGlnProHisAsnAsnArg----- 280
Db 1241 TGAAGGCTGGGTGTATTGGCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1300
QY 281 -----PheMetile-- 283
Db 1301 ATGAGGACAGTGTACTCTGATCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1360
QY 284 -CysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaAr 303
Db 1361 CTCTCTGTACAGTGTCTTCTT-----GAAGTCTCTGCTATCAGCCAGGAG 1408
QY 303 gGlyArgLeuLeuGluArgAsnGly-GluAspTyrlleCysProAsnCysThrIleLeuG 323
Db 1409 AGGA-----CACAGGAGGAGGAGGAGAT----- 1431
QY 323 lnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLys----- 339
Db 1432 -----GAGGACAGTGATACAGGATCAGCCGAGGATGAAGAGAGGAGTGTGCTT 1486
QY 340 -----TrpArgProGly---AspAlaA 346
Db 1487 CCACACCCCTGCAAGTGTCTTCTTGAAGGCTTGGGTGTATCGGCCAGGAGGAGGAGGAGGAG 1546

```

```

QY 346 spGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyI 366
Db 1547 AGGAGGAGGAGATGAGGATGTGGATAGT---GAGGATAAGGAGAGATGATTTCAGAGCAG 1603
QY 366 leLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheG 386
Db 1604 CCTTAGGAGAGCTGAGTCAGACCCACATCCTCC----- 1638
QY 386 lnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHisValA 406
Db 1638 ----- 1638
QY 406 laGlnProAspSer-ValTyrcysSer-----AsnAspCysIleLeuLysHisAlaAla 423
Db 1639 --CACCCGAGCAGAGTGTCCCATCTTCAGAGGCTGGGATATCGACCTGTAAGAGAGAGCAG 1696
QY 424 AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMet 443
Db 1697 AGGAAGAGAGAGTGTCTGAGGAGCTGGGAGAGAGCTGAGCCCTTCCGAGTGGCCA 1756
QY 444 LysMetLysProGluLys-----ProSerLeu 452
Db 1757 TCTATGTACTCTGGAGAGAGAGCCACCGCTCCTCGGCTCTCTCTAGGCTGCCCTCCGAC 1816
QY 453 ProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaPro 472
Db 1817 TCAAGAGGAGGCTCAAGCGCCAGAACCCCTACTCATGATCGGAGCCCTGAGACTCCCC 1876
QY 473 GluLysLysGlu-----ThrThrValLysLysAlaValVal----- 485
Db 1877 TAAAGGCCAGAAAGTGGCTTCTCCGAGAGGTCATCTGTCTCCTTCTCTGCTGTCTGG 1936
QY 486 -----ProAlaArg-----SerGluAlaLeuGlyLysGluAla 496
Db 1937 CAGGCGCGGCCAGGCGCCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1996
QY 497 AlaCysGluSerSerThrProSerTrpAlaSerAspHisAsnTyraAsnAlaValLysPro 516
Db 1997 GCTTCGACAGCGCATCACCC-----AGGCC 2023
QY 517 GluLysThrAlaAlaProSerProSerLeuLeu 527
Db 2024 AGGAGGAGCTGAGCCCTTCTCCTCACCCCTGTCTG 2056

RESULT 15
US-09-105-537-30/c
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Alignment Scores:
Pred. No.: 0.00465 Length: 13842
Score: 158.50 Matches: 143
Percent Similarity: 32.62% Conservative: 71
Beat Local Similarity: 21.80% Mismatches: 223
Query Match: 5.30% Indels: 221
DB: 3 Gaps: 33

```


OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 15:46:03 ; Search time 3266 Seconds
(without alignments)
5138.563 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPDSNEAPKAIKPTS.....RSFWTAIPWACPLGVAALC 562

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DRV=xlp
-Q/cgn2/USPTO.spool_p/US09787016/runat 30042004 083702 4981/app query.fasta_1.711
-DB=EST -Qfmt=fastap -SUFFIX=pen.rst -MINMATCH=0.1 -LOOPEXT=0 -COOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CEN 1 4237 @runat 30042004 083702 4981 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DET_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gsa1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2976	99.6	2752	11	BC014615 Homo sapi
2	2114.5	70.7	4619	11	AK042474 Mus muscu
3	2100.5	70.3	4651	11	BC060249 Mus muscu
4	1752.5	58.6	1640	11	BC029110 Mus muscu
5	1689	56.5	1614	11	AK014422 Mus muscu
6	1557.5	52.1	1086	9	AL578222
7	1403.5	47.0	925	9	AL546121 AL546121
8	1362	45.6	891	13	EX350413 EX350413
9	1300.5	43.5	1018	9	AL550675 AL550675
10	1275.5	42.7	1093	9	AL555810 AL555810
11	1225	41.0	850	14	CD364822 UI-R-FT2-
12	1180.5	39.5	773	12	BG545979 602573278
13	1156	38.7	1033	9	AL580669 AL580669
14	1154.5	38.6	1036	13	BO881380 AGENCOURT
15	1144.5	38.3	760	10	BE298017 601118492
16	1127	37.7	653	10	AW373892 QV0-BT053
17	1127	37.7	653	10	AW373929 QV3-BT053
18	1104.5	37.0	772	10	BE299127 601118879
19	1101	36.8	1201	9	AL514138 AL514138
20	1099	36.8	868	10	BF306721 601888796
21	1089.5	36.5	874	13	BU522670 AGENCOURT
22	1081	36.2	781	13	BQ221479 AGENCOURT
23	1074	35.9	923	10	BF308603 601890469
24	1043	34.9	1201	9	AL553387 AL553387
25	1041.5	34.8	898	13	BQ646878 AGENCOURT
26	1035	34.6	650	10	BE298593 601119753
27	1027.5	34.4	1119	9	AL550431 AL550431
28	1020	34.1	740	13	BU689908 UI-CF-FNO
29	1005	33.6	1050	13	EX404270 EX404270
30	1002.5	33.5	3385	11	AK044919 Mus muscu
31	973	32.6	692	10	BE393471 601310142
32	963.5	32.2	653	10	BF308654 601890233
33	948	31.7	660	10	AW960420 EST372491
34	944	31.6	635	13	BQ639998 he22h10.Y
35	942.5	31.5	1201	13	EX442517 EX442517
36	941	31.5	751	10	BF525689 602070017
37	935	31.3	871	13	BU131981 603116323
38	928.5	31.1	670	13	BU204261 604154862
39	920	30.8	842	13	BQ437271 AGENCOURT
40	919	30.7	510	10	BE293103 601106174
41	910.5	30.5	740	12	BI913500 603179127
42	899	30.1	752	13	BU111948 603127356
43	890.5	29.8	583	12	BM018416 603646110
44	888.5	29.7	866	13	BQ212019 AGENCOURT
45	864	28.9	512	10	BE245092 TCHAP1D25

ALIGNMENTS

RESULT 1
BC014615
LOCUS Homo sapiens cDNA clone IMAGE:3532298, containing frame-shift
DEFINITION BC014615 2752 bp mRNA linear HTC 17-DEC-2003
ACCESSION BC014615
VERSION BC014615.2 GI:33869935
KEYWORDS errors
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2752)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenman, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heien, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McEwen, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, D.J., Rulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2752)
Strausberg, R.
Direct Submission
Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:15779097.
Contact: MGC help desk
Email: cgapsb@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 11 Row: 4 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18375616
This clone has the following problem: frame shifted.

FEATURES

source

1..2752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3532298"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH MGC 17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Alignment Scores:
Pred. No.: 1.03e-189 Length: 2752
Score: 2976.00 Matches: 562
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 0
Query Match: 99.57% Indels: 1
DB: 11 Gaps: 0

US-09-787-016A-3 (1-562) x BC014615 (1-2752)

QY 1 MetAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20

DB 230 ATGGACGACCAAGGCGACCGAGCATGAGGCGACCTAAGGCCATCAACCCACCAGC 289

QY

21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40

DB 290 AARAGATTTCAGGAACACATGGGGTTTTCGAAGGACCACCTATATCCCAAGCGAGGGCGCA 349

QY

41 GlyAspAlaGluAlaAspProLeuGluProProProProGlnGlnGlnGlnGlyLeuSer 60

DB 350 GGGGACCGGAGGCTGACCCACTGGAGCGCCACCACCCACAGCAGCAGCTGGGCTCTGCC 409

QY

61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80

DB 410 CTGGGGCGAGTGGAGGCGAGCCACAGCGCACTGGAGCGCTGGAGCAGTTCTCTGACCAT 469

QY

81 AlaArgArgGlyArgArgSerMetProValSerLeuGlnAspSerGlyGluProThr 100

DB 470 GCGCGGCGCGCGCGAGGAGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 529

QY

101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120

DB 530 TCTTGGCCCCCGACAGCGCGAGCAGCTCTCGAGGCGAGCGCTGGAGGCGCTTCTGAG 589

QY

121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140

DB 590 ACCAGAGCGCGCGCGCGAGTCTGCTTCCACAGCTGTGAAGGAGACGACGCTCTTCTGAA 649

QY

141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160

DB 650 AAGGTAAAGAGGAGGAGTACCGATGACACCTCCGATGATGACCGATGGCTGACC 709

QY

161 LeuLysGluLeuGlnAsnArgLeuArgGluGlnGlnGlnGlnGlnGlnGlnGlnGln 180

DB 710 TTGAACAGCTTCAGATCGCTTCGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 769

QY

181 LeuLysGlyLeuGlnSerArgLeuArgGlyValArgGluGlnGluGlyProAlaGluThr 200

DB 770 CTGAAGGGATCCAGACTCGCTCGGAGAGAGCGCGGAGGAGGCTCCGCCGAGACT 829

QY

201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220

DB 830 GTGGGCTCCGAGGCGGAGTGTGAGGAGGCGCTCTGCGCCAGTACAGCAGGAGCGCG 889

QY

221 AsnAspGlnGlyValSerGlnAlaGlyValAspAspArgGluSerLysLeuGluGly 240

DB 890 AACGATCAGGGGTTGTGTCCAGGCTGGGAGAAAGATGACAGAGAGATGAAGTTGGAG 949

QY

241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260

DB 950 AAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGACTTGGCGCGAGCGGAGCGTGA 1009

QY

261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnArg 280

DB 1010 TGTGAGGTTACGACCCCAACGCTCTGTTATTTGCAATTTGCCCGCCAGCTTCACAAACAGG 1069

QY

281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyLeuSer 300

DB 1070 TTTATGATTTCTGTGACCGCTGTGAAGATGGTTTCATGGCGATTGTGTGGSCATTCT 1129

QY

301 GluAlaArgGlyArgLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320

DB 1130 GAGGCTCGAGGCGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1189

QY

321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340

DB 1190 ATTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGGCGATCAGCAGAGACTAAATGG 1249

QY

341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlnLysSer 360

DB 1250 AGACCTGGAGATGCTGATGGCACCAGTTGTACAAGTATAGGAAACAATAGAGCAGAAGTCT 1309

QY

361 SerGluAspGlnGlyIleLysGlyArgIleGlnLysAlaAlaAsnProSerGlyLysLys 380

DB 1310 AGCGAAGCAAGGAGTAAGGGTAGAATTGAGAAGCTGCATAATCCCAAGTGGCAGAG 1369

TDVETASRGVSSESRISGVPVSLGKEHPASSEKAKGCEBEEDTSDSDGLTIKE
 LQNRLLREDEPERSLQSONRLRKKREEDSARTSGVQISGAEQDPUCKQPEA
 SQGPVSQSEDDIENQLEKATQNTENPREAGKPECEVYDFPNADYICROPHNN
 RFMCCDRSEDFHGDYVIGSEARGLLERNGEDYICNPCTILOQDETNGSATBQD
 SGRGVSADGDTDSIGTIEQSDQIGRIEKAANPSKXKCLKI PQYVEAPGAP
 KCIQPGSVAPDSVQSCNDILKHAATWRLSSGKEQTKPKRKVKTKPKFSLP
 KCSVOVGKI SSVHRLASERENPVKKVMLASSETSKKAACESSTPSWASDHNTN
 AVKPEKPTALSLKCTYHPKAGPPGSHLGCLGLSTRVLGVLVLIVASS
 SLPAERSRQDASGPVFLPSLWSLGNFLKSCVGLMLBAISYFSFRPW*

ORIGIN

Alignment Scores:

Pred. No.: 1,85e-131 Length: 4619
 Score: 2114.50 Matches: 423
 Percent Similarity: 81.22% Conservative: 31
 Best Local Similarity: 75.67% Mismatches: 95
 Query Match: 70.74% Indels: 11
 DB: 11 Gaps: 7

US-09-787-016A-3 (1-562) x AK042474 (1-4619)

QY 1 MetAspArgGlyAspProSerLeuGluAlaProLysAlaIleProThrSer 20
 DB 299 ATGGATGATAAAGGCGCCTGAGCATGAGGAGGACCCACAGCTATCAAAACCCACAGT 358
 QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaIleArgGluGlyAla 40
 DB 359 AAGGAGTTCAGGAAACCTGGGGTTTCGAGAGACCCAGATTGCCAAAGCTGAGGGTGA 418
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGluGlnGlnLeuGlyLeuSer 60
 DB 419 GAGAGACAGGAGTGGACCCAGTGGAGCAGCAACCA-----CAGCAGCATTAACCTCTCC 472
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 473 CTGCGCGGAGTGACCGGACCAACCAACCTACTGAGAGGGTAGAGAGTTCTTACCAG 532
 QY 81 AlaArgArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 533 GTTCGGCGCGCAGGGAAGAAAGATGCGCGGTCTCCCTGGAGGATTCCAGTGGAGGCCA 592
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 593 TCTTCACAGTCACTGATGTGGAGACAGCTTCCAGGGGAGCGTTGAAAGCAGTCTGAG 652
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
 DB 653 ATCAGAAAGTGGCCCTGTATCTGACTCTCTAGGG---AAGAACATCTCTGCCCTCTGAA 709
 QY 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 710 AAGGCAAAAGGAGGTGAGAGGAGGAGAGACCTCTGACAGTGCAGTGTATGCGCTTACG 769
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
 DB 770 TTGAAGGAACTTCAGAACCCCTTCGGAGAGCGAGAGCAACCTGTGGAGAGTCC 829
 QY 181 LeuLysGlyLleGlnSerArgLeuArgLysLysArgGluGluGlyProAlaGluThr 200
 DB 830 CTGAGAGGCGAGTCAAGATCGCTCGAGAGAGAGCGAGCGAGAGAGATCTTCCCGAAAT 889
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
 DB 890 ---GGAGTGTCCAAATAGGAGTGGCGGAGCAGCAGACAGCTCTCTGTAAAGCAGAGGCT 946
 QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
 DB 947 GAGGCTAGTCAGGCGACAGTGTCTCCAGTCAAGACAGATGACATAGAAAATCAGTTGGAA 1006
 QY 240 GlyLysAlaAlaGlnAspLleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
 DB 1007 GGGAGGCGAGTCAAGGAATACAGAGGAAACCCAGGAGGAGCGGCAACCAAGCTT 1066

QY 260 GluCysGluGlyTyArgProAsnAlaLeuTyCysIleCysArgGlnProHisAsn 279
 DB 1067 GAGTGTGAGGTTTACGACCCCAATGCCCCCTGTACTGCATCTGCCCCAGCCTCACACAC 1126
 QY 280 ArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyLe 299
 DB 1127 AGGTTTATGATCTGCTGTGATGGGTGTGAGGAGTGGTTCCATGCTGCTGTGTGGTATT 1186
 QY 300 SerGluAlaArgGlyArgLeuLeuGluArgGlnGlyGluAspTyIleCysProAsnCys 319
 DB 1187 TCTGAGGCCCGAGGCGGCTCTCTGAAAGGAACCGGGAAGACTACATCTGCCCAATTCG 1246
 QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLys 339
 DB 1247 ACCATTTTGCAGTGCAGGATGACACAAACGGTAGGCGCCATGAGCAGACTCTCGG 1306
 QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
 DB 1307 TGCAGATCTGTGGTGTCTGATGGCACACACTGCACACATAGGGAACAATAGAGCAGAAG 1366
 QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
 DB 1367 TCCGAGAGAGACCGAGGCATTAAGGTTAGGATTCGAGAGGCGACAAACCCAGCGGCAAG 1426
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 DB 1427 AAAAACTCAAGATATTCAGCCTCTGCTAGAGGCTCTGCTGCTCTCTCTCTCTCTCTCT 1486
 QY 400 ProGlyCysCysHisValAlaGlnProAsnSerValTyCysSerAsnAspCysIleLeu 419
 DB 1497 CTGTGGTGTTCAGTGTAGACACGCTGACTCTGTGTATTCAGTATGATGATGATGATGAT 1546
 QY 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
 DB 1547 AAACACGACGAGCTTACCATGAGATTTCTAAGTTTCAAGTAAAGAACAAACAAACCC 1606
 QY 440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
 DB 1607 AAGGAAAAAGTCAAGACGAGGCGAGAAAGTTTCAAGTCTTCCAAATTCAGTGTTCAGTG 1666
 QY 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrVal 479
 DB 1667 GCGATTAAATCTCTTCTGTGCACAGAGACTAGCCTCAGAGAAAGGAGGAAACCCAGTG 1726
 QY 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
 DB 1727 AAGAAA---GTGATGCTGGCTTCCAGGAGTGAGACTTCTGGGAGGAGGAGCCTGTGAG 1783
 QY 500 SerSerThrProSerThrAlaSerAspHisAsnTyAsnAlaValLysProGluLysThr 519
 DB 1784 AGCAGCACACCATCTCTGGCAAGTGACCACTACAACTGCTGTGAGAGCGAGAGGCCA 1843
 QY 520 AlaAlaPro-----SerProSerLeuLeuTyLysCysMetTyHisLeuGlyVal 536
 DB 1844 GAGAAAGCCCATGCACTCTGCCCCACCTTATGTGATTAATGTACGTATCACCAAGGCT 1903
 QY 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
 DB 1904 GGCCTCCAGGCGCCCTCCATC--ATCTGGGTGGCT---GCTGGGGCTGTCTAGGA 1955
 RESULT 3
 BC060249 4651 bp mRNA linear HTC 19-NOV-2003
 LOCUS Mus musculus death inducer-oblierator 1, mRNA (cDNA clone
 IMAGE:5826043), containing frame-shift errors.
 DEFINITION
 ACCESSION BC060249
 VERSION BC060249.1 GI:37805411
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4651)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, K.H., Shermen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heish, F., Diatchenko, L., Marusina, K., Parmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, A.F., Casavant, T.L., Carnetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Arinanci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Heiton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A.G., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smaluk, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 4651)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (20-Oct-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgaps-rc@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smaluk, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Maranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 126 Row: k Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES

source
1. .4651
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6826043"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_lib="NIH BMAP_FY0"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

ORIGIN

Alignment Scores:
Pred. No.: 1.64e-130 Length: 4651
Score: 2100.50 Matches: 422
Percent Similarity: 81.07% Conservative: 32
Best Local Similarity: 75.36% Mismatches: 95

Query Match: 70.27% Indels: 12
DB: 11 Gaps: 7
US-09-787-016A-3 (1-562) x BC060249 (1-4651)
QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaLysProThrSer 20
DB 278 ATGGATGATAAAGGGGACCTGAGCAGTATGAGGAGCACCAGCGCTATCAACCCAGT 337
QY 21 LysGluPheArgLysThrTyrGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
DB 338 AAGGAGTTCAAGGAAACCTGGGGTTTCGAGACACACGATTCGCCAAGCGTGGGTGCA 397
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 398 GGAGACACGAGGTGAGACCCAGTCAGCAGCAACCA-----CAGCAGCATACCTCTCC 451
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 452 CTGGCCCGCAGTGAGCGGCACCAAAACGTACTGAGAGGGTAGAGAGTTCTTACCAG 511
QY 81 AlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 512 GTTCGGCGCGGAGGAAAGAAATGTCGGGTGCTCTGGAGATTCAGTCAGTCCACCA 571
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 572 TCTTCCACAGTCAGTCAGTGTGAGCAGCTTCGAGGGGAGCGGTGAAAGCAGTCTTGAG 631
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 632 ATCAGAGGTGGCCCTCTATCTGCTCTTACCTTACCTCTCTCTCTCTCTCTCTCT 688
QY 141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 689 AAGGCAAAAGAGAGGTGAGAGGAGAGAGACCTCTCGACAGTCAGCAGTGTGCGCTTACG 748
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
DB 749 TTGAGGAGACTTCAGACCGCTTCGAGAAAGCAGAGACAGACCTCTGTGAGAGGTCC 808
QY 181 LeuLysGlyLysGlnSerArgLeuArgLysLysArgGluGluGluGlyProAlaGluThr 200
DB 809 CTGAGAGGAGTCAGAAATCGCTCGAGGAGAGAGCGCAGAGAGAGAGATTCGCGCAACT 868
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
DB 869 ---GGAGGTGTCAAATAGCGAGTCCGCGAGCAGACACCTCTCTGTAGCAGAGGCTT 925
QY 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGlnSerLysLeuGlu 239
DB 926 GAGGCTAGTTCAGGACACCTGTCCTCAGTCCAGACAGACATGACATAGAAATCAGTTGGAA 985
QY 240 GlyLysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
DB 986 GGGAGGGCGACTCAGGGAAATACAGAGGAAACCCCGAGGAGCGGCGCAACCAAGCTT 1045
QY 260 GluCysGluGlyLysAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
DB 1046 GAGTGTGAGTTTACGACCCCAATGCCCTGTACTCATCTGCGCGCAGCTCACAACAAC 1105
QY 280 ArgPheMetIleCysCysAspArgCysGlnGluThrPheHisGlyAspCysValGlyLys 299
DB 1106 AGSTTTATGATCTGCTGTGATCGGTGTGAGGAGTGTTCATGTGGTGTGTGTGGGTATT 1165
QY 300 SerGluAlaArgGlyArgLeuLeuGlnArgAsnGlyGluAspTyrIleCysProAsnCys 319
DB 1166 TCTGAGGCCCGAGGGCGGCTCTCGAAGAGACCGGGAGAGACTACATCTGCCCAATTGC 1225
QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
DB 1226 ACCATTTTGCAGTGCAGATGAGACAAACCGTAGCGCACCGCAGCAGTCTCTGGG 1285

QY 340 TptArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnlys 359
 Db 1286 TGCAGATCTGGGGTCTGATGGCAGACACTGCACAGTAGGACAGTAGGACAGAG 1345
 QY 360 SerSerGluAspGlnGlyIleGlyGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
 Db 1346 TCCGAGAGACCCAGGGCATAAAGGTTAGGATTGAGAGGCGAGCAAAACCCAGCGGCAAG 1405
 QY 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGly 399
 Db 1406 AAAAACAACCTCAAGATATTCAGCCCTGCTGTAGAGGCTCTGTGTCTTAAATGATGATGGC 1465
 QY 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 419
 Db 1466 CTGGGTCTTCAGGTAGACAGCGCTGACTCTGTGTATTGCGATGATGACTGCTTC 1525
 QY 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPr 439
 Db 1526 AAACACGAGCAGCTACCATGAGATTTCTTAAGTTTCAGGTAAAGAACAAAAAACAAAC 1585
 QY 439 GlyGlyLysMetLysMetLysProGlyLysProSerLeuProLysCysGlyAlaGlnAl 459
 Db 1586 CAAGGAAAGGTCAGACGACGACGAAAGTTTCAGTTCTTCAAAATGCACTGTTTCAGGT 1645
 QY 459 aGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrVal 479
 Db 1646 GGGGATTAATAATCTCTTCTGTGCACAGAGACTAGCTTCAGGTTCAGAGAAAGGAAACCCAGT 1705
 QY 479 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysG 499
 Db 1706 GAAGAAA---GATGCTGGCTTCAGGAGTGAGACTTCTGGAGAGGAGGAGCTGTGGA 1762
 QY 499 uSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysTh 519
 Db 1763 GAGCAGCACACCATCTGGGCAAGTGACCAACATACATCTGTGAAGCCAGAGAGCC 1822
 QY 519 rAlaAlaPro-----SerProSerLeuLysLysCysMetTyrHisLeuGlyVa 536
 Db 1823 AGAGAAGCCCACTGCTCCCTCCCTATGTAGTAATGTAGTATCATCCCAAGGCC 1882
 QY 536 lGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
 Db 1863 TGGCTTCCAGGCGCCCTCCCATC-ATCTGGGTGCT--GCTGGGCTGTCTAGGA 1935
 RESULT 4
 BC029110
 LOCUS
 DEFINITION Mus musculus death inducer-obliterator 1, mRNA (cDNA clone
 IMAGE:3980050), containing frame-shift errors.
 ACCESSION BC029110.1 GI:21137689
 VERSION
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1640)
 REFERENCES
 AUTHORS Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
 Carrinci, P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalon, D.K., Muzny, D.J., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: coapbe-femail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
 Series: IRAC Plate: 23 Row: k Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein
 This clone has the following
 Location/Qualifiers

1..1640
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3980050"
 /tissue_type="Mammary tumor. Metallothionien-tgF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI CGAP_Mam1"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 1,13e-107 Length: 1640
 Score: 1752.50 Matches: 343
 Percent Similarity: 83.83% Conservative: 25
 Best Local Similarity: 78.13% Mismatches: 66
 Query Match: 58.63% Indels: 5
 DB: 11 Gaps: 4
 US-09-787-016A-3 (1-562) x BC029110 (1-1640)
 QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
 Db 332 ATGGATGATTAAGGGCACCTGAGCAATGAGAGACCCCAAGGCTATCAACCCACCACT 391
 QY 21 LysGluPheArgLysThrTptGlyPheArgThrThrIleAlaLysArgGluGlyVala 40
 Db 392 AAGAGTTTCAGGAAACCTGGGTTTCGAAGAACCAAGATTGCCAAACGTTGAGGGTGA 451
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 Db 452 GGAGACACGGAGGTGGACCCCGAGTCAGCAGCAACCA-----CAGCAGCATTAACCTTCC 505
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80

Db 506 CTGCGCGGAGTGGAGCGGCAACCAACGCTACTGAGAGGTAGAGGTTTCTTACCAG 565
QY 81 AAlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
LOCUS 566 GTTCGGCGCGGAGGAAAGAAATGTCGGTGTCTCTGAGGATTCAGTGGAGCCACA 625
Db 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
QY 626 TCTTCCACAGTCACTGATGTGGAGAGAGCTTCCGAGGGAGCGGTGAAGAGCAGTTCTGAG 685
Db 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
QY 586 ATCAGAGTGGCCCTGTATCTGACTCTTAGGG--AAAGAACATCTGCTCTTCTGAA 742
Db 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
QY 743 AAGCGAAAGAGAGTGAAGGAGGAGAGACACCTCTGACAGTGCAGTGTGAGCCCTTACG 802
Db 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
QY 803 TTGAGGAATTCAGAACCCCTTCGGAGAAAGCGAGGACAGACCTGTGGAGAGGTCC 862
Db 181 LeuLysGlyGlnSerArgLeuArgLysArgArgGluGluGluProAlaGluThr 200
QY 863 CTGAGAGGCACTCAGATCCCTCAGGAGAGAGCGAGAGGAGATTCGTGCGCAACT 922
Db 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro--SerLysGlnGluPro 219
QY 923 ---GGGAGTGTCAAATGAGCGTCCGAGCAGGACAGACCTCTCTGTAGCAGGAGCCT 979
Db 220 GluAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
QY 980 GAGGCTAGTCAGGAGCCAGTGTCCAGTCAGGAGAGACAGTACATAGATAAAATCAGTTGAA 1039
Db 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
QY 1040 GGGAGGCGACTCAGGGAATATACAGAGAAACCCAGGAGCGGCAACCAAGAGCCT 1099
Db 260 GluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
QY 1100 GAGTGTGAGGTTTACGACCCCAATGCCCTGTACTGTACTGTCCATCTGCCCGACCTCAACAAC 1159
Db 280 ArgPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIle 299
QY 1160 AGGTTTATGATCTGCTGTGATCGGTGTGAGAGTGTGTCCATGTGACTGTGTGGGTATT 1219
Db 300 SerGluAlaArgGlyArgLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
QY 1220 TCTGAGCCCGAGGCGGCTCTCTGGAAGGAAACGGGAGAGACTACATCTGCCCAATTCG 1279
Db 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
QY 1280 ACCATTTTCAAGTGCAGGATGAGCAAAACGCTAGCGCCACCGATGAGCAGGACTCTGGG 1339
Db 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
QY 1340 TGCAAGTCTGTGGTGTGTGATGGCAGCAGACTGCAACAGTAGGAGCAGTAGGAGCAGAG 1399
Db 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379
QY 1400 TCCGGAAGACAGGCGGCAATAAGGGTAGGATTGAGAAGGAGCAGCAAAACCCAGCGGCAAG 1459
Db 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysValleGly 399
QY 1460 AAAAAAATCAGATATTCCAGCCCTCTGTAGAGGCTCTGTGGTCTCTTAATGCAATTGGC 1519
Db 400 ProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeu 419
QY 1520 CTTGGGTGTTCAGTGTAGCAGACCTGACTCTGTGATTGCAATGACTGCAATTCCTC 1579
Db 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 438

Db 1580 AAACAGCGAGCAGCTACCATGAGATTCTTAAGTTCAGGTAAAGAACAAAAACAAA 1636
RESULT 5
LOCUS AK014422 1614 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830408B01 product:death inducer-obliterat-1, full insert sequence.
ACCESSION AK014422
VERSION AK014422.1 GI:12852263
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
6 (bases 1 to 1614)
7 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hizamoto, K., Hirooka, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, E., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp)

Eukaryota; Metazoa; Chordata; Vertebrata; Ruteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1086)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 16, 2001 this sequence version replaced gi:12942093.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

FEATURES

ORIGIN

Alignment Scores:

Alignment Scores:						
Pred. No.:	8,81e-95	Length:	1086			
Score:	1557.50	Matches:	232			
Percent Similarity:	97.34%	Conservative:	1			
Best Local Similarity:	97.01%	Mismatches:	5			
Query Match:	92.11%	Indels:	4			
DB:	5	Gaps:	2			

TS-09-787-016A-3 (1-562) x AL578222 (1-1086)

263	QY	GlyTyrAspProAsnAlaLeuTyrCysIle---CysArgGlnProHisAsnAsnArgPhe	281
1067	Db	GGTTAGACCAACGGCCC-----TGTATGGTGTGCCCCAGGCTCAACAACAG-TTT	1015
282	QY	MetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGlu	301
1014	Db	ATGATTGCTGTGACCCCTGTGAAGAATGGTTTCATGGCGATTGTGTGGCATTTCTGAG	955
302	QY	AlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIle	321
954	Db	GCTCGAGGGAGGCTTTTGGAAAGGAATGGGAAGACTATATCTGCCCAACTGCACCATT	895
322	QY	LeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArg	341
894	Db	CTCAAGTGTGAGATGAGACTCATATTCAGAAACGGCGATCAGCAGGAAGCTAAATGGAGA	835
342	QY	ProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSer	361
834	Db	CCTGGAGATGCTGATGGCACCAGTTGTACAAATATAGGAACAATAGACAGAGTCTAGC	775
362	QY	GluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLys	381
774	Db	GAAGACCAAGGATAAAGGGTAGAATTGAGAAAGCTGCAATCCAAATGGTGCAGAAGAARA	715
382	QY	LeuTyrIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGly	401
714	Db	CTCAAGATCTTCAGCCTGTGATAGAGGCCCTGGTGCCTCAAAATGTATGTGCCCGGG	655
402	QY	CysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHis	421

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 1.65e-84 Length: 925
 Score: 1403.50 Matches: 276
 Percent Similarity: 97.53% Conservative: 0
 Best Local Similarity: 97.53% Mismatches: 7
 Query Match: 46.96% Indels: 5
 DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x AL546121 (1-925)

QY 155 AspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgLysArgGluGln 174
 DB 70 GACAGCATGCGCTGACCTTGAAGAGCTTCAGATCGCTTCGCGAGGAGCGGAAACAG 129
 QY 175 GluProThrGluArgProLeuLysGlyLeuGlnSerArgLeuArgLysArgGluGln 194
 DB 130 GAGCCACCTGAGGCGCCCTGAAGAGNATCCAGAGTGGCTCGGAGAGGCGCGGAG 189
 QY 195 GluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuPro 214
 DB 190 GAGGTCCTCCGCGAGACTGTGGCTCCGAGGCGGAGTGCACCTGTGGAGGCGTCTCTGCC 249
 QY 215 SerLysGlnGluProGluAsnAspGlnGlyValValSerGluAlaGlyLysAspArg 234
 DB 250 ACTAAGCAGGAGCGGAGACCATCAGGCGGTTGTCTCCAGCTGGGAAAGATGACAGA 309
 QY 235 GluSerLysLeuGluGlyLysAlaGlnAspLysLeuAspGluGluProGlyAspLeu 254
 DB 310 GAGAGTAACTGGAGGAAAGCGCGCTCAGACATCAAGATGAGGAGCTGGAGACTTG 369
 QY 255 GlyArgProLysProGluCysGluGlyTyrrAspProAsnAlaLeuTyrrCysIleCysArg 274
 DB 370 GCGCAGCGAGGCTGAACTGAGGCTTACGACCCCAAGCGCTGATTCATTCCTGCC 429
 QY 275 GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluLysPheHisGly 294
 DB 430 CAGCTTCACAAACAGGTTTATGATTTGCTGTGACCGCTGTGAAGATGGTTTATGCC 489
 QY 295 AspCysValGlyLysSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrr 314
 DB 490 GATTGTGTGGCACTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGATGCGAAGACTAT 549
 QY 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
 DB 550 ATCTGCCCAACTGCACCATTTCTGCAAGTGCAGGATGAGACTCATTCAGAAAACGCGAGT 609
 QY 335 GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
 DB 610 CAGCAGAGCTTAATGGAGACTGAGATGCTGATGGCACCATTTGTACAGTATAGGA 669
 QY 355 ThrIleGluGlnLysSerGluAspGlnGlyLysGlyArgIleGluLysAlaAla 374
 DB 670 ACAATAGACAGAGTCTAGCGAAGACCAAGGGATTAAGGATGAGATTGAGAAAGCTGCA 729
 QY 375 AsnProSerGlyLysLysLysLysLysIlePheGlnProValIleGluAlaProGlyAla 394
 DB 730 ATCCAAAGTGGCAGAGAAAGAAATCAAGATCTTCAGCCTGTGTATAGAGGCGCTGTGTC 789
 QY 395 SerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyrrCysSer 414
 DB 790 TCAAAATGATTTGGCCCGGCTGTCTCAGCTGGCGCA-CCGACCTCGGTGTACTGCACT 848
 QY 415 AsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerGlyLysGlu 434
 DB 849 NATGACTGTAT-CTCAACACCGCGGAG-ACAATGAAGTTCTA-AGCTCAGGTAARGA- 904
 QY 435 GlnLysPro 437

DB 905 CAGAAGCCA 913

RESULT 8

BX350413/c

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. (Bases 1 to 891)

Li, W. B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seq@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 9799.1 For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0BAI036ZE02_CS03373_1&cluster=9799.1.

Contact: Peng liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID: CS0BAI036ZE02_CS03373_1.

Location/Qualifiers

1. 891

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK02J13"

/cell_type="HELA"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 9.77e-82 Length: 891

Score: 1362.00 Matches: 272

Percent Similarity: 93.28% Conservative: 6

Best Local Similarity: 91.28% Mismatches: 19

Query Match: 51.57% Indels: 4

DB: 13 Gaps: 1

US-09-787-016A-3 (1-562) x BX350413 (1-891)

QY 24 ArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAlaGlyAspAla 43

DB 890 AGGAAACCA---GGGTTTCGAGAGGCCACTATCGCCAG-CAGAGGGGCCCGAGACCG 835

QY 44 GluAlaAspProLeuGluProProGlnGlnGlnLeuGlyLeuSerLeuArgArg 63

DB 834 NAGNCTGACCCACTGAGCGCGCACCCACAGCAGCAGCTGGGCTGTCTTNCGGGCG 775

QY 64 SerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIleAlaArg 83

DB 774 AGTGAAGACCAAGCCCAAGCGACTGAGCGGTGGAGCAGTTCCTCACCACATCGCGGCG 715

QY 84 ArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThrSerCysPro 103

DB 714 GCGGCAAGNAGACATCGCTGTCTCCCTENAGGATTCCTGTGACCAACCTCTCTGCC 655

QY 104 AlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSer 123

```
Db 654 GCCACAGACGCCGAGACAGCCTCGAGGGGAGCGTGGAAAGCGCTTCTGAGACCGAGAGC 595
Qy 124 GlyProGlnSerAlaSerThrAlaValLysGluValGlyProAlaSerSerGluLysValLys 143
Db 594 GGCCTCCAGTCTGCTTCCACAGCTGTGAAGGACGACCGAGCTTCTGAAAGGAGGAAA 535
Qy 144 GlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGlu 163
Db 534 GAGGGGATGACACCATGACACCTCCGATAGTAGCAGCGATGCTGACCTTGAAGAG 475
Qy 164 LeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgProLeuLysGly 183
Db 474 CTTTCAGATCGCTTCGCGAGGAGCGGAGACGAGCGCCCTGAGAGGCGCTTGAAGGG 415
Qy 184 IleGlnSerArgLeuArgLysArgGluGluGluGluGluGluGluGluGluGluGluGlu 203
Db 414 ATCCAGAGTCCGCTCGGAGAGAGCCCGGAGGAGGGTCCCGCCGAGACTGTGGGCTCC 355
Qy 204 GluAlaSerAspThrValGluGluValLeuProSerLysGlnGluProGluLysGluGln 223
Db 354 GAGGCCAGTGACACTGTGAGGGCGCTCTGCCCGATGAGCGAGCGCCGAGAACGATCAG 295
Qy 224 GlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaLa 243
Db 294 GGGGTTGTGTCCTCCAGGCTGGGAAGATGACAGAGAGAGTAGTTGGAGGAGAGCGGCT 235
Qy 244 GlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGluCysGluGly 263
Db 234 CAGGACATCAACATGAGGAGCGCTGGAGACTTGGGCGGACCGAAGCGCTGATGTGAGGGT 175
Qy 264 TyrAspProAsnAlaLeuThrCysIleCysArgGluProHisAsnAsnArgPheMetIle 283
Db 174 TAGACCCCCAACCCCTGTATTGATTTGCGCGCAGCGCTCACACACACAGGTTTATGATT 115
Qy 284 CysCysAspArgCysGluGluTTPpHeHisGlyAspCysValGlyIleSerGluAlaArg 303
Db 114 TGCTGTGACCGCTGTGAAGATGCTTTCATGGCGATGTGTGGGCACTTCTGAGGCTCGA 55
Qy 304 GlyArgLeuLeuGluArgAsnGlyGluAspThrIleCysProAsnCysThrIle 321
Db 54 GGGAGGCT-TTGGAAAGGAATGGAA-GACTATATCTGCCCAACTGCACCAT 3

RESULT 9 /c
AL550675/c 1018 bp mRNA linear EST 31-MAY-2003
LOCUS AL550675 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1056YK04 3-PRIME, mRNA sequence.
ACCESSION AL550675
VERSION AL550675.2 GI:31272492
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Bukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
AUTHORS Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12887874.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1056BF02NP1&cluster=9799.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0D1056BF02NP1.
FEATURES
Location/Qualifiers
1..1018
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1056YK04"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores:
Pred. No.: 1.59e-77 Length: 1018
Score: 1300.50 Matches: 249
Percent Similarity: 93.96% Conservative: 0
Best Local Similarity: 93.96% Mismatches: 15
Query Match: 43.51% Indels: 2
DB: 9 Gaps: 1

US-09-787-016A-3 (1-562) x AL550675 (1-1018)
Qy 298 GlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysPro 317
Db 908 GGCATTTCTGAGCTCGAGGGA---GCTTTTGAAGGAATGGGA-GACTATATCTGCCCCA 853
Qy 318 AsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlu 337
Db 852 AACTGCACCATTTCTCAAGTCGAGGATGAGACTCAATTCAGANACCGCAGATCAGCAGAA 793
Qy 338 AlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlu 357
Db 792 GCTAATGAGAGACCTCGAGATGCTGATGCACCGATTGTACATATAGACACATAGAG 733
Qy 358 GlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSer 377
Db 732 CAGAGTCTMGCGGAGACCAAGCGATTAAGGCTGGAATTCAGAAAGCTGCAATCCAACT 673
Qy 378 GlyLysLysLysLysLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCys 397
Db 672 GGCAGAGATATCTCAAGATCTTCAGCGCTGTGATAGAGCGCTGCTGCTTCTCAATGT 613
Qy 398 IleGlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCys 417
Db 612 ATTGCCCCCGGTGCTGTCACTGCGGCGAGCCGACTCGTGTGTACTGAGTAACTACTGT 553
Qy 418 IleLeuLysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysPro 437
Db 552 ATCTCTCAAACACGCCGCGACCAATGAAGTTCTTAAGCTCAGGTAAAGACAGAGCCCA 493
Qy 438 LysProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAla 457
Db 492 AAGCCTTAAGMAAAGATGAGATGAGCCAGAGAGCCCGACTTCCGAAATGCGGTGCT 433
Qy 458 GlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThr 477
Db 432 CAGGCAGGTATTAAATCTCTTCTGTGCAAGAGACCCAGCTCCAGAAAAAAGAGAGACC 373
Qy 478 ThrValLysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAla 497
Db 372 ACAGTGAAGAGAGAGTGTGTGCTCCCTGCGGAGTGAAGCACTCCGAGAGAGAGTGTGCT 313
Qy 498 CysGluSerSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGlu 517
Db 312 TGTGAGAGCAGCAGCGCGCTGCTGGGCGAGCGATCAATTAATGACATTAAGCCAGAA 253
Qy 518 LysThrAlaAlaProSerProSerLeuLeuThrLysCysMetTyrHisLeuGlyValGly 537
Db 252 AAGACTGTGCTCCCTCGCGCTGCTGTTGTATTAATGATGATATACCTTAGGGGTGGC 193
Qy 538 LeuLeuAspProSerArgSerPheIleAlaIleProTrpAlaCysProGlyLeuGly 557
Db 192 CTCTCGAGACCCCTCCCGTCTTCTGTGATAGCATCCCTCGGCGCTGTCCAGACTGGGA 133
```



```

/db_xref="taxon:9606"
/clone="UI-H-FT2-bjm-d-20-0-UI"
/tissue_type="Aveolar Macrophage"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FT2"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP FT2 is a subcloned cDNA library constructed from
a pool of 81 RNA samples from Alveolar Macrophages
challenged with different treatments. The library was
subtracted according to Bonaldo, Lennon and Soares, Genome
Research, 6:791-806, 1996. The tissue was provided by Dr.
Gary W. Hunninghake of the University of Iowa.
TAG RTSSUE=Human Lung Aveolar Macrophage
TAG_LIB=UI-H-FT2
TAG_SEQ=GCCCATGCCG"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,52e-72 Length: 850
Score: 1225.00 Matches: 229
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 2
Query Match: 40.98% Indels: 0
DB: 14 Gaps: 0

US-09-787-016A-3 (1-562) x CD364822 (1-850)

QY 332 ThrAlaaspGlnGluAlaLysTrrpGlyAspAlaaspGlyThrAspCysThr 351
DB 850 ACGGAGATCAGCAGGAGGCTAATGAGACCTCGAGATCTGATGCCCGGATGTACA 791
QY 352 SerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyIleGlu 371
DB 790 AGTATAGACATAGACGAGAGTCTAGCGAAGACCAAGGATTAAGGCTAGATTGAG 731
QY 372 LysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLysLysLysLysLys 391
DB 730 AAAGCTGCANATCCAGTGCNAAAGAAAGAACTCAAGATCTTCCAGCGCTGTGTAGAGCG 671
QY 392 ProGlyValSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerVal 411
DB 670 CTTGGTCCCTCAAAATGTTTGGCCCGGGTGTGTCTACGTGGCGCAGCCGAGCTCGGTG 611
QY 412 TyrCysSerAsnAspCysIleLeuLysHisAlaAlaAlaThrMetLysPheLeuSerSer 431
DB 610 TACTGCAGTAATGACTATCTCTCAACACGCGCCGACGACCAATGAAGTTTCTAAGCTCA 551
QY 432 GlyLysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSer 451
DB 550 GGTAAAGACAGAGCCCAAGCCCTAAAGAAAGATGAAGTGAAGCAGAGAGCCGAGT 491
QY 452 LeuProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysAsgProAla 471
DB 490 CTTCCGAAATGCGGTGCTCAGGCAGGTATTAAATCTCTTCTGTGCACAGAGACCAAGCT 431
QY 472 ProGluLysLysGluThrThrValLysLysAlaValValProAlaAsgSerGluAla 491
DB 430 CCGAATAAAGAGACACACAGTCAAGAGGACGTGGTGGTCCCTGCGCGAGTGAAGCA 371
QY 492 LeuGlyLysGluAlaAlaCysGluSerSerThrProSerThrAlaSerAspHisAsnTyr 511
DB 370 CTCGGAGAGAGGAGCTTGTGAGAGCAGCACGCGTGGTGGGCGGAGCCATCAATTAC 311
QY 512 AsnAlaValLysProGluLysThrAlaAlaProSerProSerLysLeuLysCysMet 531
DB 310 AATCAGTAAAGCCAGAAAGACTGTCTCTCCCTCGCGGTCTACCTGTGTATTAATGTATG 251
QY 532 TyrHisLeuGlyValGlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrp 551
DB 250 TATCACCTAGGGGTGGCTCTCTGAGACCCCTCCCGTCTTCTGATAGGATCCCTCGG 191

```

```

QY 552 AlaCysProGlyLeuGlyValAlaAlaLeuCys 562
DB 190 GCGTGTCCAGGACTGGAGATTGCAGCTTTGTGT 158

```

RESULT 12

```

BG545979
LOCUS
DEFINITION
602573278F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4701475 5',
mRNA sequence.
ACCESSION
BG545979
VERSION
BG545979.1 GI:13544644
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 773)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgepbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM1536 row: 1 column: 20
High quality sequence stop: 699.
FEATURES
Location/Qualifiers
1..773
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4701475"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattagcc); Site 2: SfiI (ggccattagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

```

FEATURES

```

source
1..773
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4701475"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattagcc); Site 2: SfiI (ggccattagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,33e-69 Length: 773
Score: 1180.50 Matches: 234
Percent Similarity: 92.28% Conservative: 5
Best Local Similarity: 90.35% Mismatches: 12
Query Match: 39.49% Indels: 8
DB: 12 Gaps: 1

US-09-787-016A-3 (1-562) x BG545979 (1-773)

QY 238 LeuGluGlyLysAlaAlaGlnAspIleLysAspGluGlnProGlyAspLeuGlyArgPro 257
DB 2 TTGGGAGGAAAGCGCGCTCAGGACATCAAGATGAGAGCGCTGGAGACTTGGCCCGGACCG 61
QY 258 LysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHis 277
DB 62 AAGCCTGAATGTGAGGGTTACGACCCACAGCCCTGTATTGCACTTTCGCCAGCCTCAC 121
QY 278 AsnAsnArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysVal 297
DB 122 AACACAGGTTTATGATTTCTGTGACCGCTGTGAAGAATGGTTTCTATGCGCATTTGTGTG 181

```


QY 298 GlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspThrIleCysPro 317
Db 182 GGCATTTCCTGAGGCTCGAGGGAGGCTTTTGGAAAGCAATGGGAGACATATATCTGCCCA 241
QY 318 AsnCysThrIleLeuGluValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGlu 337
Db 242 AACTCACAATCTCTCAAGTGCAGATGAGATCTATTCAAGAACGGCGAGATCAGCAGAA 301
QY 338 AlatySerArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGlu 357
Db 302 GCTAAATGGAGACCTGGAGATCTGATGCCCGATTGTACAAAGTATAGGAACAATAGAG 361
QY 358 GlnIleSerSerGluAspGlnGlyIle-LysGlyArgIleGluIleValAlaAlaAsnProSe 377
Db 362 CAGAAAGTCTAGCGAAGACCAAGGATACAGGGTAGATTGAAAGCTGCAATCCAG 421
QY 377 rGlyLysLysLysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCys 397
Db 422 TGGCAAGAAGAACTCAAGATCTTCAGCCTGTGATAGAGGCGCTGTGCTCAAAATG 481
QY 397 sIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCy 417
Db 482 TATTGGCCCGGGTGTGTGACGTGGCGAGCCGACTCGGTGTACTGATGATGATG 541
QY 417 sIleLeuLysHisAlaAla-AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysP 437
Db 542 TATCTCAACACGCGCGCAGGACATGAAGTTCTTAAGCTCAGTAAAGAACAGAACG 601
QY 437 rGlyProLysGluLysMet-LysMetLysProGlu-LysProSerLysProLysCys 455
Db 602 CAAGCGTTTACAGAAACGATGAGATGGAAGCCAGAGAAAGCCAGTCTTCGAAATGC 661
QY 456 GlyAla-GlnAlaGlyIle-LysIleSerSerValHisLysArgProAlaProGluLysL 475
Db 662 GGGGCTCCAGGCGGGTATTCACCACTCTCTCTGTGCAACAAGAGACCAAGTCCAGAAAG 721
QY 475 ysGluThrThr-ValLysLysAla-ValValValProAlaArg 488
Db 722 AAGAGAACACAGGTGAAGAAGCCAGGGGTGGTTCCTCGGCGC 764

RESULT 13
AL580669
LOCUS
DEFINITION
Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
AL580669
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Homo sapiens
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1033)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:12946913.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9799.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ007YD06
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DJ007YD03QP1.

FEATURES
source
1..1033
/organism="Homo sapiens"
/mol_type="mRNA"

/db xref="taxon:9606"
/clone="CS0DJ007YD06"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
Pred. No.: 8,57e-68 Length: 1033
Score: 1156.00 Matches: 231
Percent Similarity: 96.69% Conservative: 3
Best Local Similarity: 95.45% Mismatches: 8
Query Match: 38.68% Indels: 1
DB: 9 Gaps: 0

US-09-787-016A-3 (1-562) x AL580669 (1-1033)

QY 1 MetAspAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
Db 304 ATGACCGCAAGGCGAATGAGGAGGACCTAAGGCATCAACACCCACGAGC 363
QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db 364 AARAGTTTCAGAAATCATGGGTTTTCGAGGACCACTATCGCCAAAGCGAGGCGCA 423
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 424 GGGACCGGAGGCTGACCCACTGGAGCGCCACCCACAGCAGCAGCTGGGCTGTCC 483
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
Db 484 CTGGGCGCAGTGGGAGGACGCCCAAGCGCACTGAGCGCGTGGAGCAGTCTCTGACCAT 543
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 544 GCGCGCGCGCGCGAGGAGCATCTCTTCTCCCTGGAGGATCTTGGTGAGGCCAG 603
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 604 TCTGCCCCCGCACAGACCGCCGAGACAGCTCCGAGGGCAGCGTGAAGAGCGCTTCTGAG 663
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 664 ACCAGAGCGGCCCCCTCTCTTCCACAGCTGTGAAGGAACGACACCGCTCTTCTGAA 723
QY 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspLysLeuThr 160
Db 724 AAGGTAAAGGAGGAGGAGTGAACAATGACACCTCCGATAGTACAGCGATGGCTGACC 783
QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Db 784 TTGAARAGCTTCAGATCGCTTCGAGGAAGCGGGAACAGGAGCCCACTGAGAGGCC 843
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGluProAlaGluThr 200
Db 844 CTGAAGGAATCCAGATGCGCTCGCGAARAGCGCCGAGGAGGCTCCCGCGAGACT 903
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 904 GTGGCTCCGAGGCCARTGACATATGAGGGCGCTCTKSCCART-AARCAAGAGCCCGAR 962
QY 221 AsnArgGlnGlyValValSerGlnAlaLysAspAspArgGluSerLysLeuGly 240
Db 963 AACGATCAAGGGGTGTGTCTCCAGGCTGGGAAAGAGAGAGAAATATGTTGAGGGA 1022
QY 241 LysAla 242
Db 1023 AAGGCG 1028

RESULT 14
BQ881380
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ881380 1036 bp mRNA linear EST 16-AUG-2002
AGENCOURT 8725620 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6340359
S' mRNA sequence.
BQ881380.1 GI:22273388
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens
Bukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1036)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-t@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM2536 row: k column: 16
High quality sequence start: 41
High quality sequence stop: 312.
High quality sequence stop: 312.

FEATURES
source
Location/Qualifiers
1..1036
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6340359"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_47"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1.09e-67 Length: 1036
Score: 1154.50 Matches: 249
Percent Similarity: 75.14% Conservative: 11
Best Local Similarity: 71.97% Mismatches: 66
Query Match: 38.62% Indels: 22
DB: 13 Gaps: 6

US-09-787-016A-3 (1-562) x BQ881380 (1-1036)

QY 172 ArgGluGlnGluProThrGluArgProLeuLysGlyLeuGlnSerArgLeuArgLysLys 191
DB 11 CGGTCTACTCGGGACTGAGAGGCCCTTACAGA-ATACAGAGTGGCTGCGGAGAGAG 69
QY 192 ArgArgGluGluGluProAlaGluThrValGlySerGluAlaSerAspThrValGluGly 211
DB 70 CGCGGGAGGAGGGTCCCGCGAGACTGTGGCTCCGAGGCCAGTGCACCTGTGGAGGGC 129
QY 212 ValLeuProSerLysGlnGluProGluAsnAspGlnValValSerGlnAlaGlyLys 231
DB 130 GTCTGCTCCAGTACGAGGAGCCCGAGACGATCAGGGGGTGTGTCTCCAGGCTGGGAA 189
QY 232 AspAspArgGluSerLysLeuGluGlyValAlaGlnAspLysLeuGluPro 251

DB 190 GATGACAGAGAGTAAGTTGGAGGGAAGCGGCTCAGGACATCAAGATGAGGAGCCT 249
QY 252 GlyAspLeuGlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCys 271
DB 250 GGAGACTTGGGCGGACCGAAGCCTGAATGTGAGGGTTAGACCCCAACCCCTTGA-TGT 308
QY 272 IleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluTTP 291
DB 309 TTTTGGCCGACCTCACCAACAGGTTATGATTTGCTGTGACCGCTGTGAAGATGG 368
QY 292 PheHisGlyAspCysValGlyIleSerGluAlaAa-glyArgLeuLeuGluArgAsnGly 311
DB 369 TTTTATGGGATTTGTGTGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGATGG 428
QY 312 GluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGlu 331
DB 429 GAAGACTATATCTGCCCAACTGCACCATCTGCAAGTGCAGATGAGACTTATTGAGA 488
QY 332 ThrAlaAspGlnGlnAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThr 351
DB 489 ACAGGAGATCAGCAGGAGCTAAATGGAGACTTGGAGATGCTGATGGCACCGATTGTACA 548
QY 352 SerIleGlyThrIleGlnLysSerSerGluAspGlnGlyIleLysGlyValGluGlu 371
DB 549 AGTATAGGAACTAGAGCAGAGTCTAGCAAGACCAAGGATTAAGGGGTAGATTGAG 608
QY 372 LysAlaAlaAsnProSerGlyLysLysLysLysLysLysLysLysLysLysLysLys 391
DB 609 ATAGCTGCAATCCAGTGGCAAGAACTCAAGATCTTCCAGCTTGTGTAGAGGC 668
QY 391 aProGlyAlaSerLysCysIleGlyPro-----GlyCysCysHisValAlaGlnProAs 409
DB 669 GCTGTGCTCCTCAAAATGTATTGGCCCCCGGGTGGCTGTCCACGGGGGGCGCCAGCCCCA 728
QY 409 pSerValTyrCysSerAsnAsp-----CysIleLeuLysHisAlaAlaThrMe 426
DB 729 C---CTCGGGTGTACTGGAAGTAATGACTTGGATCCCTCAAAACCGCGCGCGCGGA 785
QY 426 tLysPheLeuSerSer-----GlyLysGluGlnLysProLysProLysGlu 441
DB 786 CAAATGAAGTCTCTTAAGCCTTCGGGTAAAGGAACACGCAAGCCCGCCAGGCGCTTAAA 845
QY 441 uLysMetLysMetLysProGluLysProSer-----LeuProLysCysGlu 456
DB 846 GAAAGAGATTGAGTATTGAGAGCCCGGAGAGGCCCCCAGTCCCTTCCCTTAATGTC-- 903
QY 456 yAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysG 476
DB 904 -----GCGGGTGCTCCCGGGCAAGCGTATTAAATCCTCTTCTGTGGTGGCCCCAA 956
QY 476 uThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAl 496
DB 957 ACAAGACCCCGCTCTCTCGAGATTAACCCGACGACCCACCGCGGGAACAACAAGCGTAG 1016
QY 496 aAlaCysGluSerSer 501
DB 1017 TCGGGCGGGTCTCTCC 1032

RESULT 15
BQ881380
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BQ881380 760 bp mRNA linear EST 20-JUL-2000
601118492F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3028271 5',
mRNA sequence.
BQ881380.1 GI:9181603
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 760)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 1, 2004, 14:32:06 ; Search time 554 Seconds
(without alignments)
4309.544 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989
Sequence: 1 MDDKGPSEBAPKAIKPTS.....RSFWIAPNACPLGVAALC 562

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q=/cgn2_1/USFIO_spool_p/US09787016/runat 30042004 083701 4961/app query fasta_1.711
-DB=N Geneseq 29Jan04 -OFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016 @CGN 1 1 708 @runat 30042004 083701 4961 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2989	100.0	2610	3 AAZ90578	Aaz90578 Human dea
2	2981	99.7	2772	4 AAH15102	Aah15102 Human cdn
3	2742	91.7	2085	5 ABV28029	Abv28029 Human pro
4	2742	91.7	2085	5 ABV27935	Abv27935 Human pro
5	2742	91.7	2085	5 ABV22095	Abv22095 Human pro
6	2737	91.6	2085	5 ABV22192	Abv22192 Human pro
7	2449	81.9	7838	4 AAK52368	Aak52368 Human pol
8	2440	81.6	7290	4 AAK53352	Aak53352 Human pol

9	2262	75.7	2332	5 AAS86576	Aas86576 DNA encod
10	2112.5	70.7	2867	3 AAZ90578	Aaz90578 Murine de
11	1706	57.1	1764	3 AAC77933	Aac77933 Human can
12	1467	49.1	2986	5 AAS86575	Aas86575 DNA encod
13	989	33.1	30626	4 AAK67051	Aak67051 Human imm
14	898	30.0	503	5 AAS86574	Aas86574 DNA encod
15	807	27.0	478	8 ACH31915	Ach31915 Human end
16	785	26.3	481	5 ABV42461	Abv42461 Human pro
17	561.5	18.8	362	5 ABV12528	Abv12528 Human pro
18	549	18.4	681	4 AAH08005	Aah08005 Human cdn
19	527	17.6	365	5 ABV42575	Abv42575 Human pro
20	527	17.6	365	5 ABV33664	Abv33664 Human pro
21	494	16.5	324	6 ABR21248	Abn21248 Human ORF
22	471	15.8	408	3 AAK43780	Aak43780 Mouse sec
23	447	15.0	7838	4 AAK52368	Aak52368 Human pol
24	420	14.1	256	5 ABV03359	Abv03359 Human pro
25	329.5	11.0	6051	4 ABL29341	Ab129341 Drosophil
26	311	10.4	7174	9 ADC37602	Adc37602 Human nuc
27	267.5	8.9	8751	4 ABL29340	Ab129340 Drosophil
28	246	8.2	636	5 ABV03229	Abv03229 Human pro
29	232.5	7.8	2387	3 AAF22374	Aaf22374 Human sec
30	232	7.8	5975	6 ABL61775	Ab161775 Colon ade
31	232	7.8	6256	4 AAI58390	Aai58390 Human pol
32	232	7.8	6256	8 ADB48359	Adb48359 Novel hum
33	232	7.8	6948	5 AAH81806	Aah81806 Human dif
34	232	7.8	7354	4 AAK52041	Aak52041 Human pol
35	226	7.6	6584	4 AAI60176	Aai60176 Human pol
36	226	7.6	6996	4 AAK53025	Aak53025 Human pol
37	221.5	7.4	6947	7 ACF34533	Acf34533 Gene enco
38	219.5	7.3	593	5 ABV12398	Abv12398 Human pro
39	217	7.3	1992	4 ABL27413	Ab127413 Drosophil
40	214.5	7.2	2320	5 ABX71405	Abx71405 Human tes
41	212.5	7.1	1161	4 ABL27411	Ab127411 Drosophil
42	203.5	6.8	10314	4 ABL29797	Ab129797 Drosophil
43	203	6.8	659	3 AAC08973	Aac08973 Human sec
44	203	6.8	915	4 AAS26292	Aas26292 Human cdn
45	203	6.8	915	7 ABX73633	Abx73633 Human nov

ALIGNMENTS

RESULT 1

AAZ90578
ID AAZ90578 standard; cDNA; 2610 BP.

XX AC AAZ90578;
XX DT 19-JUN-2000 (first entry)
XX DE Human death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; human;
XX KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
XX KW antirheumatic; antiinflammatory; antiproliferative; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX CDS 265..1953
XX FT /*tag= a
XX FT /product= "DIO-1"
XX PN WC2000015787-AL.
XX PD 23-MAR-2000.

XX 10-SEP-1999; 99WO-GB003019.

XX 10-SEP-1998; 98SE-00003069.

XX 17-SEP-1998; 98US-0100873P.

XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
XX (BANN/) BANNERMAN D G.

Db 1885 CCCTCCGGTCTTCTTGATAGCCATCCCTGGGCTGTCCAGGACTGGAGTTGCAGCT 1944
QY 561 LeuCys 562
Db 1945 TTGTGT 1950

RESULT 2
AAH15102
ID AAH15102 standard; cDNA; 2772 BP.
AC AAH15102;
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:13129.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPT; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX Claim 8; SEQ ID NO 13129; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX polynucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 2772 BP; 699 A; 691 C; 765 G; 617 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.26e-176 Length: 2772

Score: 2981.00 Matches: 561
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.73% Indels: 0
DB: 4 Gaps: 0

US-09-787-016A-3 (1-562) x AAH15102 (1-2772)

QY 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db 301 ATGGACGACAAAGGGGACCCGAGCAATGAGAGGACCTTAAGGCCATCAACCCACGAC 360
QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrIleAlaLysArgGluGlyVala 40
Db 361 AAAGAGTTCCAGGAAACATGGGGTTTTTCGAAGACCACTATCCCAAGGAGAGGGCGCA 420
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 421 GGGGACGGGAGGCTGACCCACTGGGAGCCGCCACCCACAGCAGCAGCTGGGCTGTCC 480
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheIleuThrIle 80
Db 481 CTGCGGCGCAGTGGGAGGAGCCCAAGCGCACTGAGCGGCTGGGAGCAGTTCTTGACCAT 540
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 541 GCGCGCGCGCGCGGAGGAGGAGCATGCTGTCTCTCCCTGGAGGATCTCTGTGAGCCACG 600
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 601 TCCTGCCGCCGCCACAGACGCCGAGCAGCGCTCCGAGGCGCAGCGTGGAAAGCGCTTTCGAG 660
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db 661 ACCAGAGGCGGCGGCGGAGGAGTCTGCTTCCACAGCTGTGAGGAGACCAACGAGCTCTTCTGAA 720
QY 141 LysValLysGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Db 721 AAGCTGAAGAGGCGGATGACCACTGACCACTCCGATAGTCAGCAGCGATGCGCTGAC 780
QY 161 LeuLysGluLeuGlnAsnArgLeuArgGlyArgGluArgGluGlnGluProThrGluArgPro 180
Db 781 TTGAAGAGGCTTCAGAAATCGCTTCGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
QY 181 LeuLysGlyIleGlnSerArgLeuArgLysArgGluGluGluGlyProAlaGluThr 200
Db 841 CTGAAGGAGATCCAGAGTCTGCTCGGAGAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 900
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db 901 GTGGGCTCCGAGGCGGAGTGCACCTGTGGAGGGGCTCTGCGCCAGTAAAGCAGGAGCCGAG 960
QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Db 961 AACGATCAGGGGGTGTGTTCCAGGCTGGGAGAGATGACAGAGAGAGTAAAGTTGGAGGGA 1020
QY 241 LysAlaAlaGlnAspLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db 1021 AAGCGGCTCAGGACATCAAGATGAGGAGGAGCTGAAGACTTGGCGCGAGCAAGAGCTGAA 1080
QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
Db 1081 TGTGAGGTTTACGACCCCAAGCCCTGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 1140
QY 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
Db 1141 TTTATGATTGCTGTGAGCGCTGTGAAGAATGTTTTCATGGCGGATTTGTGGGCAATTTCT 1200
QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyValAspThrIleCysProAsnCysThr 320
Db 1201 GAGCTCGAGGAGGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAACTGCAC 1260
QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340

Db	1261	ATTCTCAAGTGCAGGATGAGACTCTATTAGAAACGGCAGATCAGCGGAAGCTAAATGG	1320
Qy	341	AtcProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnIlysser	360
Db	1321	AGACCTGGAGATGCTGATGGCACCGATTGTACAAGTATAGGAACAATAGACGAGAAGTCT	1380
Qy	361	SerGluAspGlnGlyIleIysGlyArgIleGluIysAlaAlaAsnProSerGlyIlyssys	380
Db	1381	AGCGAAGACCAAGGGATAAAGGGGTAGAAATTGAGAAAGCTGCAATCCAAGTGGCAGAG	1440
Qy	381	LysLeuIlystIlePheGlnProValIleGluAlaProGlyValaserIysCysIleGlyPro	400
Db	1441	AAACTCAAGATCTTCAGCCTGTGATAGAGCGCCCTGGTCCCTCAAAATGATATGGCCCC	1500
Qy	401	GlyCysCysHisValAlaGlnProAspSerValtyrCysSerAsnAspCysIleLeuIys	420
Db	1501	GGGTGCTGTCACGTGGCGCAGCCCGACTCGGTGTACTGCAGTAATGCTGTATCTCTCAA	1560
Qy	421	HisAlaAlaIleThrMetIlyssPheLeuSerSerGlyIysGluGlnIlyssProIys	440
Db	1561	CACGCCGCGACGCAATGAAGTTTCTAAGCTCAGGTAAAGAACGGAAGCCAAAGCCTAAA	1620
Qy	441	GluIysMetIlyssMetIlyssProGluIlyssProSerLeuProIlyssCysGlyAlaGlnAlaGly	460
Db	1621	GAAGAAGTGAAGATGAGCCAGAGAGGCCAGCTTCCGAAATGCGGTGCTCAGGCAGGT	1680
Qy	461	IleIlyssIleSerSerValHisIlyssArgProAlaProGluIlyssGluThrThrValIys	480
Db	1681	ATTAAATCTCTTCTGTGTGCAAGAGACCCAGCTCCAGAAAAAAGAGAGCCACAGTGAAG	1740
Qy	481	LysAlaValValProAlaArgSerGluAlaLeuGlyIlyssGluAlaAlaCysGluSer	500
Db	1741	AGGCAGTGTGTGTCTCTGTGCGGAGTGAAGCACTCGGGAAGGACGAGCTTGTGAGGC	1800
Qy	501	SerThrProSerTrpAlaSerAspHisAsnTyraSnAlaValIlyssProGluIlyssThrAla	520
Db	1801	AGCACGCCGTCGTGGGCGAGCGATCACAATACAAATGCAGTAAAGCCGAGAAAGRCTGCT	1860
Qy	521	AlaProSerProSerLeuLeuTyrllyssCysMetTyrlHisLeuGlyValGlyLeuLeuAsp	540
Db	1861	GCTCCCTCGCGCTCAGCTGTGTATTAATGTATGTATCACTAGGGGTTCGCTCTCTGGAC	1920
Qy	541	ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyIleuGlyValIlaAla	560
Db	1921	CCCTCCCGTTCTTCTCGATAGCCATCCCTCGGCGCTGTCAGGACTGGGAGTTCAGCT	1980
Qy	561	LeuCys 562	
Db	1981	TTGTGT 1986	
RESULT 3			
ABV28029			
ID	ABV28029 standard; cDNA; 2085 BP.		
XX			
AC	ABV28029;		
XX			
DT	16-SEP-2002 (first entry)		
XX			
XX	Human prostate expression marker cDNA 28020.		
XX			
XX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;		
KW	pharmacogenomic marker; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	WO200160860-A2.		
XX			
PD	23-AUG-2001.		
XX			
PF	20-FEB-2001; 2001WO-US005171.		
XX			
PR	17-FEB-2000; 2000US-0183319P.		

PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

PS
PS Claim 1; Page 5788; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX ;
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 7,23e-162 Length: 2085
Score: 2742.00 Matches: 551
Percent Similarity: 98.39% Conservative: 0
Best Local Similarity: 98.39% Mismatches: 9
Query Match: 91.74% Indels: 3
DB: 5 Gaps: 0

US-09-787-016A-3 (1-562) x ABV28029 (1-2085)

QY 1 MetAspAlpsGlyAspProSerAsnGluAlaProLysAlaLeuLysProThrSer 20
Db 276 ATGGACACAAAGCGCCGCCGATGAGGAGCCTTAGGCATCAACACCACGAGC 335
QY 21 LysGluPheArgLysTrpTgLyPheArgThrThrileAlaLysArgGluGlyAla 40
Db 336 AAAGAGTTTCAGGAANAACATGGGTGTTTCGAAGGACCATATCGCCACGAGCGGCGCA 395
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db 396 GGGGACCGGAGCGTGACCCACTGGAGCCGCA -CCCCCACAGCAGCAGCTGGGCTGTCC 454
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrle 80
Db 455 CTGCGGCGCATGGGAGGCGACGCCAACAGCGCACTGAGCGCGTGGAGCAGTTCTCGACCAT 514
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db 515 GCGCGGCGCGCGCAGGAGAGCATGCTGTCTCTCCCTGGAGGATTCTGGTGAGCCGACG 574
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db 575 TCCTGCCGCCG -ACAAGAAGCCGACACAGC -TCCGAGGCGACGCTGGAAGAGCGCTTCTGAG 632
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaVallysGluArgProAlaSerSerGlu 140
Db 633 ACCAGAAGCGGCCCCAG -TCTGCTTCACAGT -GTGAAGGAACGACCCAGC -TCITCTGAA 689
QY 141 LysVallysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160

690 AAGGTGAAGAGGGGATGACCAAGATGACACCTCGATAGTGACAGCGATGGCTG-ACC 748
161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGluGluProThrGluArgPro 180
749 TTTAAAGAGCTTCAGAAATCGCTT-CGCAGAGAGCGGAAACGAGGCCCACTCAGAGGCC 807
181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGluGlyProAlaGluThr 200
808 CTGAAGAGGATCAGAGTGCCTCGGAGAGAGCGCGGAGGAGGGTCCCGCCGAGACT 867
201 ValGlySerGluAlaSerAspThrValGluGluValLeuProSerLysGlnGluProGlu 220
868 GTGGCTCCGAGGCCAGTGCACACTGTGGAGGGCGTCTGCCCACTAAGCAGGAGCCGAG 927
221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgLysSerLysLeuGluGly 240
928 AACGATCAGGGGGTGTGTCCCGAGGTGGGAAAGATCAGAGAGGTAAAGTTGGAGGA 987
241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
988 AAGGCGGCTCAGGACATCAAGATGAGGAGCGCTGGAGACTTGGGCCGACCGAAGCCTGAA 1047
261 CysGluGlyThrAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArg 280
1048 TGTGAGGGTACACCCCAACGCCCTGTATTGCTTTCGCGCAGCTTCAACAACAGG 1107
281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
1108 TTTATGATTTGCTGACCCCTGTGAGATGGTTTCATGGCGATTTGTGGCAATTTCT 1167
301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTrpIleCysProAsnCysThr 320
1168 GAGGCTCGAGGGAGGCTTTTGGAAAGAAATGGGGAAGACTATATCTGCCCAACTGCACC 1227
321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLysTrp 340
1228 ATTCTGCAAGTCAGAGTATGATCAATTCAGAAACCGCAGATCAGCAGAGCTAAATGG 1287
341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
1288 AGACCTGGAGATCTGATGGCACCGGATTGTACAGTATAGCAATATAGCAGCAAGATCT 1347
361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
1348 AGCGAAGACCAAGGATAAAGGGTAGAATTTGAGAAAGCTCAATCCAGTTGCCAAGAG 1407
381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
1408 AAATCAGATCTTCAGCTCTGATAGAGCGCGCTGCTCCCTCAAAATGATTGGCCCC 1467
401 GlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLys 420
1468 GGGTGTCTGTCACGTGGCGCAGCGGACTCGTGTACTGCAAGTAAGACTGTATCCCAAA 1527
421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
1528 CACGCCACGCGACAAATGAGTTTCTAAGCTCAGTAAAGACAGAGGCCAAAGCCTAAA 1587
441 GluLysMetLysMetLysProGluLysProSerLysProLysCysGlyAlaGlnAlaGly 460
1588 GAAAGATGAAGATGAAGCCAGAGAGCCGAGTCTTCCGAAATGCGGTCTCAGGCGAGGT 1647
461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
1648 ATTAATAATCTTCTGTGTCACAGAGACCGAGCTCCAGAAAAAAGAGAGCAGTGAAG 1707
481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
1708 AAGCAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGGAGGAGGAGCTGTGTGAGAGC 1767
501 SerThrProSerTrpAlaSerAspHisAsnThrAlaValLysProGluLysThrAla 520
1768 AGCAGCGCGTGTGGGGAGGAGATCAATATCAATGATGATGAGGAGGAGAGAGCTGCT 1827

521 AlaProSerProSerLeuLeuTyLysCysMetTyHisLysLeuGlyValGlyLeuLeuAsp 540
1828 GCTCCCTCGCGTCACTGTGTATATAATGTATATCACTAGGGTTGGCTCTCTGCAC 1887
541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
1888 CCTCCCGTCTTCTTCTGGATAGC-ATCCCTGGGCGCTGTCCAGGACTGGAGTTGAGCT 1946

RESULT 4
ABV27935
ID ABV27935 standard; cDNA; 2085 BP.
XX
AC ABV27935;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 27926.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US0051171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5757-5758; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 7,23e-162 Length: 2085
Score: 2742.00 Matches: 551
Percent Similarity: 98.39% Conservative: 0
Best Local Similarity: 98.39% Mismatches: 9
Query Match: 91.74% Indels: 9
DB: 5 Gaps: 0

US-09-787-016A-3 (1-562) x ABV27935 (1-2085)

1 MetAspLysGlyValSerProSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
Db ATGACGCAAAAGCGACCCCGCAATCAGAGGGCACTTAAGGCCATCAAAACCCACCCAGC 335
21 LysGluPheArgLysThrTyrGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
Db AAAGAGTTTCAGGAAACATGGGTTCGAAAGGACCACTATCGCCAGCGAGGGCGCA 395
41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
Db GGGGACGGAGGAGTGAACCACTGAGCGCGCA-CCCCCAGAGCAGCAGCTGGGCTGTCC 454
61 LeuArgSerGlyArgGlnProLysArgThrGluArgValGluInPheLeuThrIle 80
Db CTGCGCGCAGTGGAGGAGCAGCCCAAGCGCACTGAGCGCTGGAGCAGTTCCTGACCAAT 514
81 AlaArgArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
Db GCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 574
101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Db TCTCTCCCGCGC-ACAGACGCCGAGACAGC-TCCGAGGCGCAGCGTGGAAAGCGCTTCTGAG 632
121 ThrArgSerGlyProGlnSerThrAlaValLysGluArgProAlaSerSerGlu 140
Db ACCAGAGCGGCGCCCGAG-TCTGCTTCCACAGT-GTGAAGGAGACGACCCAGC-TCCTTCTGAA 689
141 LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
Db AAGGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 748
161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
Db TTGAAGAGCTTTCAGAAATCGCTT-CGCGAGAAAGCGGGAACAGAGGCCCACTGAGAGGCC 807
181 LeuLysGlyIleGlnSerArgLeuArgLysArgArgGluGluGluGlyProAlaGluThr 200
Db CTGAAGAGGATTCAGAGTCCCTGCGAAGAGAGCGCGGAGAGGGTCCCGCCGAGACT 867
201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Db GTGGGCTCCGAGGCCAGTCACACTGTGGAGGGCGCTCTGCCAGTAAGCAGAGGCCCGAG 927
221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Db AACGATCAGGGGGTGTGTCCAGGCTGGGAAGATGACAGAGAGTAAGTTGGAGGGA 987
241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Db AAGGGGCTCAGGACATCAAGATGAGGAGCCTGGAGACTTGGCGCGACCGAAGCCTGAA 1047
261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
Db TGTGAGGGTTAGACCCCAACGCCCTGTATGTGATTTGGCGCGAGCTCACACACACAG 1107
281 PheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIleSer 300
Db TTTATGATTTGCTGTGACCGCTGTGAAGATGCTTTCATGGCGATTTGTGGCGATTTCT 1167
301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
Db GAGGCTCGAGGGAGGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1227
321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
Db ATTCTGCAATGCGAGATGAGATCTATTAGAAACCGCAGATCAGCAGGAGTAATGG 1287
341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
Db AGACCTGGAGATCTGATGGCACCAGTGTGTACAAAGTATAGGAACAATAGCAGAGAGTCT 1347

361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
Db AGCGAAGACCAAGGGATAAAAGGCTAGAAATTGAGAAAGCTGCAATTCCTCAACTGGCAGAG 1407
381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
Db AAATCTCAGATCTTCCAGCTGTGATAGAGGCGCTGGTGCCTCAAAATGTATTGGCCCC 1467
401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
Db GGGTGTCTGTCACTGGCGCAGCCGCACTCGGTGTACTGCAGTAACTGTATCTCTCAAA 1527
421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
Db CACCCCGCAGGACATGAAGTTCTTAAGCTCAGGTAAGACACAGNAGCCAAAGCCTAAA 1587
441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
Db GAAAGATGATGATGAAGCCAGAGCCAGCTCTTCCGAAATCGGTGTCTCAGGCAGGT 1647
461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
Db ATTAATAATCTCTTCTGTGCAAGAGACCACTCCAGAAAAAAGAGACACACAGTGAAG 1707
481 LysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
Db AAGCAGCTGGTGTCTCTCGCGGAGTGAAGCACTCCGGAAGGAGCAGCTTGTGAGAGC 1767
501 SerThrProSerThrAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
Db AGCAGCGGTGTGTGGGCGAGCATCAATTTACATGCACTTAAGCCAGAAAAAGAGACTGCT 1827
521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
Db GCTCTCTCGCGCTCACTGTGTATAAATGTATGATATCACTAGGGGTGGCTCTCTGGAC 1887
541 ProSerArgSerPheThrIleAlaIleProThrAlaCysProGlyLeuGlyValAlaAla 560
Db CCTCTCCGCTCTTTCTGTGATAGC-ATCCCTGGGCGCTGTCCAGGACTGGAGTTGCAGCT 1946

RESULT 5
ABV22095 standard; cDNA; 2085 BP.
ID ABV22095 standard; cDNA; 2085 BP.
XX AC ABV22095;
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 22086.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US005171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 3785; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

SQ Sequence 2085 BP; 520 A; 525 C; 657 G; 378 T; 0 U; 5 Other;

Alignment Scores:
 Pred. No.: 7,23e-162 Length: 2085
 Score: 2742.00 Matches: 551
 Percent Similarity: 98.39% Conservative: 0
 Best Local Similarity: 98.39% Mismatches: 9
 Query Match: 91.74% Indels: 9
 DB: 5 Gaps: 0

US-09-787-016a-3 (1-562) x ABV22095 (1-2085)

QY 1 MetAspIysGlyAspProSerAsnGluGluAlaProIysAlaIleLysProThrSer 20
 DB 276 ATGACGACAAAGCGCCGAGCATGAGAGGACCTAAGGCCATCAAAACCCACGCG 335
 QY 21 LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla 40
 DB 336 AAGAGTTTCAGGAAACATGGGTGTTTCGAAGGACCACTATCGCCAAAGCGAGGGGCGCA 395
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 DB 396 GGGGACCGGAGGCTGACCCACTCGAGCCGCA-CCCCACAGCAGCAGCTGGGCTGTCC 454
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 455 CTGGCGGCGAGTGGAGGCGACCCAGCGCACTGAGCGCGTGGAGCAGTTCTCTGACCAT 514
 QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 515 GCGCGGCGCGCGGAGGAGCATGCTGCTCTCCCTGGAGGATTCCTGTGAGCCACG 574
 QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 575 TCTTGGCCCCGC-ACAGACGCGGAGCAGC-TCGAGGGGCGAGGTGGAAAGCGCTCTGAG 632
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 DB 633 ACCAGAAGCGCCCCAG-TCGTGCTTCACAGT-GTGAAGGAGAACACAGC-TCCTCTGAA 689
 QY 141 LysValLysGlyArgAspHisAspThrSerAspSerAspSerAspGlyLeuThr 160
 DB 690 AAGGTGAAGAGGGGATGACCAAGATGACACCTCCGATAGTACAGCATGGCTG-ACC 748
 QY 161 LeuLysGluLeuGlnAnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
 DB 749 TTGAAGAGGCTTCAGATCGCTT-CGAGGAGGCGGGAACAGGAGGCCACTCGAGAGGCC 807
 QY 181 LeuLysGlyLeuGlnSerArgLeuArgLysArgGluGluGluGluProAlaGluThr 200
 DB 808 CTGAAGAGGAGATCCAGTTCGCTCCGAGAGAGCGCCGCGAGGAGGGTCCCGCCGAGACT 867
 QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220

RESULT 6
 ABV22192

DB 868 GTGGGCTCCGAGGCCAGTCACTGTGGAGGGCGTCTCTCCCGTAGCAGGAGCCGAG 927
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgLysLeuGluGly 240
 DB 928 AAGCATCAGGGGTTCGTCTCCAGCGCTGGGAAAGATGACAGAGAGTAAGTTGGAGGGA 987
 QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
 DB 988 AAGCGGCTCAGGACATCAAGATGAGGAGCTTGGAGACTTGGCGCCGACGAGGCTGAA 1047
 QY 261 CysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArg 280
 DB 1048 TGTGAGGGTTACGACCCCAACGCGCTGTATGCTATTCGATTCGCGCCAGCTCACAAACAGG 1107
 QY 281 PheMetIleCysCysAspArgCysGluGlnTrpPheHisGlyAspCysValGlyIleSer 300
 DB 1108 TTTATGATTGCTGTGACCGCTGTGAAGATGGTTTCATGGCGAATTGTGTGGGCATTTCT 1167
 QY 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr 320
 DB 1168 GAGGCTCGAGGGAGGGCTTTTGGAAAGGATGGGAGAGACTATATCTGCCCAATGACCC 1227
 QY 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
 DB 1228 ATTCTCAAGTGCAGGATGAGACTCATTCAGAAACGCGCAGATCAGCAGGAAGCTAAATGG 1287
 QY 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 DB 1288 AGACCTGGAGATGCTGATGACCCGATTGTACAGTATATAGAAACAATAGACAGAGAGTCT 1347
 QY 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 DB 1348 AGCGAAGACCCAGGATAAAGGGTAGAATTGAGAAAGCTGCAATCCAGTGGCAAGAG 1407
 QY 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 DB 1408 AAATCAAGATCTTCAGCGCTGTGATAGAGCGCCCTGGTGCCCTCAAAATGATTTGGCCCC 1467
 QY 401 GlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys 420
 DB 1468 GGGTGTGTCTCAGTGGCGGAGCCCGACTCGGTGTACTGCAGTAATGACTGTATCTCTCAA 1527
 QY 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 DB 1528 CACGCGCAGCGCAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAGAGCCCTAAA 1587
 QY 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 DB 1588 GAAGAATGAGATGAGCAGCAGAGAGCCAGTCTTCCGAAATGCGGTGCTCAGGAGGT 1647
 QY 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 DB 1648 ATTAAATCTTCTCTGTGCACAGAGACAGCAGCTCCAGAAAAAAGAGAGACCAAGTGAAG 1707
 QY 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaLysCysGluSer 500
 DB 1708 AAGCGAGTGTGTCTCTCGCGAGTGAAGAGCTCGGAGAGAGAGAGCTTGTGAGAGC 1767
 QY 501 SerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLysThrAla 520
 DB 1768 AGCACCGCTCGTGGGCGAGCGATCAATTAATGCTAGTAAAGCCAGAAAAAGACTGCT 1827
 QY 521 AlaProSerProSerLeuLeuTyrLysCysMetTyrHisLeuGlyValGlyLeuLeuAsp 540
 DB 1828 GCTCCCTCGCGGCTACTGTGTATAATGATGATATACCTTAGGGGTGGCCCTCTGGAC 1887
 QY 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 DB 1888 CCTCTCCGTTCTTCTGCTGATAGC-ATCCCTCTGGGCTCTCCAGGACTTGGAGTTCAGCT 1946

ID	ABV22192 standard; cDNA; 2085 BP.	Db	396	GGGACGCGGAGGCTGACCCACTGGAGCCGCA - CCCCCACAGCAGCAGCTGGGGCTCTGCC	454
XX					
AC	ABV22192;	Qy	61	LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle	80
XX					
DT	13-SEP-2002 (first entry)	Db	455	CTGCGCGCAGTGGAGGCGAGCCACGCGACCTGAGGCGCTGGAGCAGTTCTGACCAATT	514
XX					
DE	Human prostate expression marker cDNA 22183.	Qy	81	AlaArgArgGlyArgArgSerMetProValSerLeuGlnuAspSerGlyGluProThr	100
XX					
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	Db	515	GCGCGGCGCGCGCGAGGAGCATGCTGCTCTCCCTGGAGGATTCTGGTGGAGCCACG	574
XX	pharmacogenomic marker; gene; ss.				
OS	Homo sapiens.	Qy	101	SerCysProAlaThrAspAlaGluThrAlaSerGlyGluSerValGluSerAlaSerGlu	120
XX					
PN	WO200160860-A2.	Db	575	TCCTGCCCCCG - ACHGACGCGGAGCAGC - TCCGAGGCGGAGCTGGAAAGCGCTTCTGAG	632
XX					
PD	23-AUG-2001.	Qy	121	ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu	140
XX					
XX	20-FEB-2001; 2001WO-US0051171.	Db	633	ACCAGAAAGCGGCCAG - TCTGCTTCCACAGT - GTGAAGGAACCCAGC - TCTTCTGAA	689
XX					
PF	17-FEB-2000; 2000US-0183319P.	Qy	141	LysValLysGlyGlyAspAspHisAspThrSerAspSerAspSerAspGlyLeuThr	160
XX					
PR	16-MAR-2000; 2000US-0189862P.	Db	690	AAGGTGAAGAGGAGGAGTACCCAGATGACACCTCCGATAGTGACAGGATGGCTG - ACC	748
XX					
PR	25-MAY-2000; 2000US-0207454P.	Qy	161	LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro	180
XX					
PR	09-JUN-2000; 2000US-0211314P.	Db	749	TTGAAAGAGCTTCAGANTGCTT - CGCAGGAAGCGGAAACAGGAGCCCACTGAGAGGCC	807
XX					
PR	18-JUL-2000; 2000US-0219007P.	Qy	181	LeuLysGlyLysGlnSerArgLeuArgLysArgGluGluGluProAlaGluThr	200
XX					
PR	13-DEC-2000; 2000US-0255281P.	Db	808	CTGAAGGAGTCCAGATGCGCTCGGAAGAGCGCGGAGGAGGGTCCCGCCAGACT	867
XX					
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	Qy	201	ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu	220
XX					
PI	Schlegel R, Endege WO, Monahan JB;	Db	868	GTGGCTCCGAGGCGCAGTGACACTGTGGAGGCGCTCTGCCACGTAAAGCAGAGCCCGAG	927
XX					
DR	WPI; 2001-662795/76.	Qy	221	AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly	240
XX					
XX	Novel isolated nucleic acid molecule associated with cancerous state of	Db	928	AACGATCAGGGGGTGTGTCCAGGCTGGGAAGATGACAGAGAGTAAGTTGAGGAGA	987
PT	prostate cells and correlating with presence of prostate cancer, useful				
PT	for detecting presence of prostate cancer, stage of prostate cancer.	Qy	241	LysAlaAlaGlnAspLysAspGluGluProGlyLysLeuGlyArgProLysProGlu	260
PS	Claim 1; Page 3816; 11750pp; English.	Db	988	AAGCGGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGGCCGACCCAGCCTGAA	1047
CC	The invention relates to an isolated nucleic acid molecule (I) comprising	Qy	261	CysGluGlyTyrAspProAsnAlaLeuTyrCysLysCysArgGlnProHisAsnAsnArg	280
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the	Db	1048	TGTGAGGTTACGACCCCAACGCGCTGTATTGCAATTTGCCGCCAGCCCTCACAACACAGG	1107
CC	specification or its complement. (I) is useful for: (a) assessing whether	Qy	281	PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer	300
CC	a patient is afflicted with prostate cancer; (b) monitoring the	Db	1108	TTTATGATTTGCTGTGCGCGCTGTGAAGATGGTTTCATGCGGATTGTGTGGCAATTTCT	1167
CC	progression of prostate cancer in a patient; (c) assessing the efficacy	Qy	301	GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThr	320
CC	of a test compound to inhibit prostate cancer in a patient; (d) assessing	Db	1168	CAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAAGACTATATCTGCCCAACTGCACC	1227
CC	the efficacy of a therapy for inhibiting prostate cancer in a patient;	Qy	321	IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp	340
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;	Db	1228	ATTTCTGCAAGTGCAGGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGCTAAATGG	1287
CC	(f) assessing the prostate cell carcinogenic potential of a compound; (g)	Qy	341	ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer	360
CC	determining whether prostate cancer has metastasized in a patient; (h)	Db	1288	AGACCTCGAGATGCTGTATGGCACCGATTTGTACAGTAAGCAACATAGAGCAGAGTCT	1347
CC	assessing the aggressiveness or indolence of prostate cancer in a patient	Qy	361	SerGluAspGlnGlyIleGlyGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys	380
CC	; (i) is also useful as a pharmacodynamic or pharmacogenomic marker	Db	1348	AGCGAAGCAAGGGATAAGGTTAGAAAGCTGCATAATCCAGTGGCAGAGAG	1407
XX	Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;	Qy	381	LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro	400
SQ		Db	1408	AAACTCAGATCTTCCAGCTGTGTAGAGGCGCTGTGTGCTCAAAATGATTATGGCCCC	1467
		Qy	401	GlyCysCysHisAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLys	420
		Db	1468	GGGTGCTGTACGTGGCGCAGCCCGACTCGGTGTACTGTCAGTAATGACTGTATCTCAAA	1527

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 3816; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 2085 BP; 520 A; 526 C; 656 G; 378 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	1,48e-161	Length:	2085
Score:	2737.00	Matches:	550
Percent Similarity:	98.21%	Conservative:	0
Best Local Similarity:	98.21%	Mismatches:	10
Query Match:	91.57%	Indels:	9
DB:	5	Gaps:	0

US-09-787-016A-3 (1-562) x ABV22192 (1-2085)

Qy	1	MetAspLysGlyAspProSerAsnGluGluAlaProLysAlaIleLysProThrSer	20
Db	276	ATGACACACAAAGCGGCCGAGCATGAGGAGCACCCTAAGCCCATCAACCCACGAC	335
Qy	21	LysGluPheArgLysThrTrpGlyPheArgArgThrIleAlaLysArgGluGlyAla	40
Db	336	AAAGAGTTTCAAGAAACATGGGGTTTTCGAAGACCACTATGCCAAGGAGAGGCGCA	395
Qy	41	GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnLeuGlyLeuSer	60

Qy 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLys 440
 Db 1528 CAGCCGACCCGACAAATGAAGTTCTTAAGCTCAGGTAAGACACAGAAAGCCCTAAA 1587
 Qy 441 GlnLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 Db 1588 GAAAGATGAAGATGAAGACGAGAGCCCAAGTCTTCGAAATCGGTCTCAGGAGGT 1647
 Qy 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 Db 1648 ATTAATATCTCTCTGTGCAACAGACCCAGCTCCAGAAAAAAGAGACACAGTGAAG 1707
 Qy 481 LysAlaValValValProAlaAlaSerGluAlaLeuGlyLysGluAlaAlaCysGluSer 500
 Db 1708 AAGSCAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGGAAGGAGCAGCTTTGTGAGAGC 1767
 Qy 501 SerThrProSerTrpAlaSerAspHisAsnTrpAsnAlaValLysProGluLysThrAla 520
 Db 1768 AGCAGCGCGCTGTGGCGAGCGATCACATTTACATGCAATGATTAAGCCAGAAAGACTGCT 1827
 Qy 521 AlaProSerProSerLeuLeuTyLysCysMetTrpHisLeuGlyValGlyLeuLeuAsp 540
 Db 1828 GCTCCCTCGCGCTCACTGTTGTATATAATGTATGTATCATCCTTAGGGGTGGCCCTCTGGAC 1887
 Qy 541 ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla 560
 Db 1888 CCCGCCCGCTCTCTGGATAGC-ATCCCTGGGCGCTGCCAGACTGGGAGTTGCGACT 1946
 RESULT 7
 AAK52368
 ID AAK52368 standard; cDNA; 7838 BP.
 AC AAK52368;
 XX
 DT 06-NOV-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 913.
 DE
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157190-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US004098.
 XX
 XX 03-FEB-2000; 2000US-00496914.
 XX 27-APR-2000; 2000US-00560875.
 XX 20-JUN-2000; 2000US-00598075.
 XX 19-JUL-2000; 2000US-00520325.
 XX 01-SEP-2000; 2000US-00549936.
 XX 15-SEP-2000; 2000US-00563561.
 XX 20-OCT-2000; 2000US-00693325.
 XX 30-NOV-2000; 2000US-00728422.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 XX WPI; 2001-476283/51.
 XX P-PSDB; AAM79235.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 XX in diagnosis and gene therapy.

PS Claim 1; Page 3007-3013; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation on which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 7838 BP; 1872 A; 2123 C; 2118 G; 1725 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.6e-143 Length: 7838
 Score: 2449.00 Matches: 477
 Percent Similarity: 90.75% Conservative: 4
 Best Local Similarity: 90.00% Mismatches: 8
 Query Match: 81.93% Indels: 41
 DB: Gaps: 4
 US-09-787-016A-3 (1-562) x AAK52368 (1-7838)
 Qy 37 ArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGln 56
 Db 528 CCGAGAGCCAGGAGAGGCT---GGGAGCG---CAGCGGCCAGACCGCGAGCGCG 581
 Qy 57 LeuGlyLeu-SerLeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluG 76
 Db 582 GGAGGCGAGCTTCCGACCT-----GGGAGCGAGCCCAAGCGCACTGAGCGCGTGGAGCA 635
 Qy 76 nPheLeuThrIleAlaArgArgGlyArgArgSerMetProValSerLeuGluAspSe 96
 Db 636 GTTCCTCACCATTGCGCGCGCGCGCGAGAGAGCATGCTCTCTCTCCCTGGAGGATTC 695
 Qy 96 rGlyGluProThrSerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValG 116
 Db 696 TGGTGAGCCCACTCTCTGCGCCGCCACAGACGCGGAGACAGCTCCGAGGCGAGCGTGA 755
 Qy 116 uSerAlaSerGluThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgPr 136
 Db 756 AAGCGCTTCTGAGACCAAGAGCGGCCCTCTGCTTCCAGAGCTGTGAAGAACGACCC 815
 Qy 136 oAlaSerSerGluLysValLysGlyLysAspAspHisAspThrSerAspSerAspSe 156
 Db 816 AGCTCTCTCTGAAAAGCTGAAGAGGGGATGACCAAGTACACCTCCGATATGACAG 875
 Qy 156 rAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluPr 176
 Db 876 CGATGGCTGACCTTGAAGAGCTTCAGAAATCGCTTTCGAGGAGAGCGGAAACAGAGGCC 935
 Qy 176 oThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluG 196
 Db 936 CACTGAGAGGCCCTCTGAAGAGGATCCAGATCTCCCTGCGAGAGAGCGCGGAGGAGGG 995
 Qy 196 yProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLy 216
 Db 996 TCCCGCGAGACTGTGGGCTCCGAGGCCAGTACACTGTGGAGGGCGCTCTCGCCCTAGTAA 1055
 Qy 216 sGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspArgGluSe 236
 Db 1056 GCAGAGGCCCGAGAACGATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGAG 1115
 Qy 236 rLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyAr 256
 Db 1116 TAGGTTGAGGGGAAAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGACTTGGGCGG 1175
 Qy 256 gProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArgGlnPr 276

QY 100 rSerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 Db 173 GTCTCTGCCCCGACAGACCGCCGAGCAGCTCCGAGGGCAGCGTGTGAAGGCTTCTCA 232
 QY 120 uThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 Db 233 GACCAGAGCGGCCCCAGTCTGCTTCCACAGCTGTGAAGGACGACGAGCTCTTCTGA 292
 QY 140 uLysValLysGlyGlyValAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuTh 160
 Db 293 AAAGGTGAAGGAGGAGGATGACACCATGACCTCCATAGTAGCAGCGATGGCCCTGAC 352
 QY 160 rLeuLysGluLeuGlnAsnArgLeuArgGlyValArgGluGlnGluProThrGluArgPr 180
 Db 353 CTTGAAGAGCTTCAGATGCGCTTCCAGAGAGCGGACAGGAGGCCCACTGAGAGGCC 412
 QY 180 oLeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluTh 200
 Db 413 CTTGAAGAGGATCCAGAGTCCCTGCGGAGAGAGCGCGGAGAGGGGTCCCGCCGAGAC 472
 QY 200 rValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 Db 473 TGTGGGCTCCGAGGCCAGTGAACCTGTGGAGGGCGTCTCTCCAGTAGCAGGAGGCCCGA 532
 QY 220 uAsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 240
 Db 533 GAACGATCAGGGGTTGTGTCCAGGCTGGGGAAGATGACAGAGAGAGTAAAGTTGGAGGG 592
 QY 240 yLysAlaAlaGlnAspIleLysAspGluGluProGlyValAspLeuGlyArgProLysProGlu 260
 Db 593 AAAGGGGGCTTCAGACATCAAGATAGGAGGCTGGAGACTTGGGCGCAGCCGAGGCTGA 652
 QY 260 uCysGluGlyTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnAr 280
 Db 653 ATGTGAGGTTACGACCCCAAGCGCTGTATTGCTATTTGCGCGCAGGCTCACAACAACAG 712
 QY 280 gPheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyIleSe 300
 Db 713 GTTTATGATTGCTGTGACCGCTGTGAAGATGGTTTCATGGCGATTTGTGGGCAFTTC 772
 QY 300 rGluAlaArgGlyValGluLeuGluArgAsnGlyGluAspTyIleCysProAsnCysTh 320
 Db 773 TGAGGCTCGAGGGAGGCTTTTGGGAAGATGGGAGAGACTATATCTGCCCAACTGCAC 832
 QY 320 rIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTr 340
 Db 833 CAITCTCAAGTGCAGGATGAGATCTCATTCAGAAACGGCAGATCAGCAGGAGGCTAAATG 892
 QY 340 pArgProGlyValAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSe 360
 Db 893 GAGACCTGGAGATGCTGATGGCCCGATTGTACAAAGTATAGGACATATAGGCGAGAGTTC 952
 QY 360 rSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLy 380
 Db 953 TAGCGAAGACCAAGGGATAAGGGTAGAATAGAAAGCTGCAATCCAAGTGGCAAGAA 1012
 QY 380 sLysLeuLysIlePheGlnPro----- 387
 Db 1013 GAAACTCAGATCTTCCAGCGCTGTCCCGGGCTGTCCCGACCCAGCTGCTGTCTCTCTG 1072
 QY 387 ----- 387
 Db 1073 GCAGGTATTGGAATTCGTGTCTAGAGAGATCTCTGCCTTCACACTCTTGCACTGCAT 1132
 QY 388 -----ValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHi 404
 Db 1133 TAGCTGCAAGGTATAGAGGGCGCTGTGTGCTTCAAAATGTATTGGCCCCGGGTGTCTGA 1192
 QY 404 sValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLysHisAlaAla 424
 Db 1193 CGTGGCGCAGCCGACCTCGGTGACTGCTAGTAACTGATGCTATCTCTCAACACGCGCAGC 1252

QY 424 aThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLy 444
 Db 1253 GACATGAAGTTTCTAAGCTCAGGTAAAGAACAGAAAGCCAAAGCCTAAAGAAAGATGAA 1312
 QY 444 sMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSe 464
 Db 1313 GATGAAGCCAGAGAAAGCCAGCTTCCCAATGCGGTGCTCAGGCAGGTATTAAATCTC 1372
 QY 464 rSerValHisLysArgProAlaProGluLysLysGluThrThrValLysLysAlaValVa 484
 Db 1373 TTCTGTGCAAGAGACCCAGCTCCAGAAAAAAGAGAGACCCAGTGAAGAGGAGTGGT 1432
 QY 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSe 504
 Db 1433 CGTCCCTGCGCGAGTGAAGCACTCGGGAAGGAGCGCTTGTGAGAGCAGCAGCCGTC 1492
 QY 504 rTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 1493 GTGGGCGAGCGATCAATATTCAGTAAAGCCAGAAAGAGACTGCTGCTCCCTCGCC 1552
 QY 524 oSerLeuLeuTyLys 529
 Db 1553 GTCACCTGTTGTATAA 1568
 RESULT 9
 AAS86576
 ID AAS86576 standard; cDNA; 2332 BP.
 XX AC AAS86576;
 XX 13-FEB-2002 (first entry)
 XX DE DNA encoding novel human diagnostic protein #22380.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX OS Homo sapiens.
 XX XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG22389.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.
 XX Claim 1; SEQ ID NO 22380; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX sequences. (I) is useful as hybridisation probes, polymerase chain
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX and in recombinant production of (II). The polynucleotides are also used
 XX in diagnostics as expressed sequence tags for identifying expressed
 XX genes. (I) is useful in gene therapy techniques to restore normal
 XX activity of (II) or to treat disease states involving (II). (II) is
 XX useful for generating antibodies against it, detecting or quantitating a
 XX polypeptide in tissue, as molecular weight markers and as a food
 XX supplement. (II) and its binding partners are useful in medical imaging
 XX of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 2332 BP; 511 A; 621 C; 764 G; 436 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.73e-132 Length: 2332
 Score: 2262.00 Matches: 477
 Percent Similarity: 79.55% Conservativity: 21
 Best Local Similarity: 76.20% Mismatches: 48
 Query Match: 75.68% Indels: 81
 DB: 5 Gaps: 12

US-09-787-016A-3 (1-562) x AAS86576 (1-2332)

QY 1 MetAspArgGlyAspProSerAsnGluAlaProLysAlaIleYsProThrSer 20
 DB 391 ATGGACGAAAGCGGACCCGAGCGATGAGAGCGACCTAAGGCCATCAACCCACGAC 450
 QY 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThrIleAlaLysArgGluGlyAla 40
 DB 451 AAAGAGTTTCAAGAAACATGGGGTTTTCGAAGGACCACTATCCCAAGCGAGAGCGCA 510
 QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnLeuGlyLeuSer 60
 DB 511 GGGACCGGAGGCTGACCCACCTGGAGCCGCCACCCCGACGAGCGCTGGCCGTCC 570
 QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
 DB 571 CTGGGGCGCATGGGAGCGGACGCCACAGCGCACTGAGCGCGCTGAGCAGATTCTCTG 630
 QY 81 AlaArgArgGlyArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 DB 631 CGCGGGCGCGCGGAGAGAGAGCTGCTCTCTCTGAGAGATTCTCTGAGCGCCAG 690
 QY 101 SerCysProLalaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 DB 691 TCTTGCCTCCGACAGCGCGGACAGACCTCGGAGCGGCGCTGGGAAAGCGCTTCTGAG 750
 QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerGlu 140
 DB 751 ACCAGAGCGGCGCCCGGCTGCTGCTTCCAGCTGTGAGGAGACGACGCTTCTTGAA 810
 QY 141 LysValLysGlyLysAspAspHisAspThrSerAspThrSerAspSerAspGlyLeuThr 160
 DB 811 AAGGTGAAGAGGAGGAGATGACCAAGATGACACCTCCGATAGTCACAGCGATGCGCTGAC 870
 QY 161 LeuLysGluLeuGlnAsnArgLeuArgArgGluGlnGlnProThrGluArgPro 180
 DB 871 TTGAAAGAGCTTCAGATCCCTTCGCGAGAGCGGGAACGAGGCGCCACTGAGAGGCC 930
 QY 181 LeuLysGlyLysGlnSerArgLeuArgLysArgGluGluGluGlyProAlaGluThr 200
 DB 931 CTGAAAGGATCCAGAGTCCCTCGGAGAGAGCGCGGAGGAGGCTCCCGCGAGACT 990
 QY 201 ValGlySerGluAlaSerArgThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 DB 991 GTGGGCTCCGAGGCGGAGTGAACCTGTGAGGCGGCTCTGCGCCAGTAGGAGGCGCGAG 1050
 QY 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGlnSerLysLeuGluGly 240
 DB 1051 AACGATCAGGGGGTGTGTCTCCCGGCTGGGAAAGATGACAGAGAGATTAAGTTGAGGCA 1110
 QY 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260

DB 1111 AAGGGCGCTCAGGACATCAAGATGAGAGCGCTTGGAGACTTTTGGCGGACCGAGCGCTGA 1170
 QY 260 uCysGluGlyTrpAspProbenAlaLeuTrpCysIleCysArgGlnProHisbenAr 280
 DB 1171 ATGTGAGGGTTTACGACCCCAACCGCTGTATTGTCAATTGGCGGCGCTCACACACAG 1230
 QY 280 gPheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyLys 300
 DB 1231 GTTATGATTGCTGTGACCGCTGTGAGATGGTTTCAAGGAGTGTGTGGCATTTTC 1290
 QY 300 rGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTrpIleCysProbenCys 320
 DB 1291 TGAGGCTCGAGGAGGCGCTTTTGGAAAGAAATGGGAGAGACTATATCTGCCCAATGCGAC 1350
 QY 320 rIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 340
 DB 1351 CATCTGCAAGTGCAGATGAGACTCATTCAGAAACGCGAGATCAGCAGAGAGCTTAAT 1410
 QY 340 rp-ArgPro-GlyAspAla-AspGlyThrAsp-CysThrSerIleGlyThrIleGluGln 358
 DB 1411 CGGAGACCTCGGAGATGCTTGTATGGCACCGGATTTTACAAAGTATAGGAAACATTAGGAGC 1470
 QY 359 Lys-----SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnPro 376
 DB 1471 AGAAGTCTTAGCCGACACCCAGGCGTTAAGGTTAGATTTGAGAAAGCTGCANATCCA 1530
 QY 377 SerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 387
 DB 1531 AGTGGCAAGAGGAACCTCAAGATCTTTCAGGCTGTATATGCGAAATGCTGTGTCTAGAAG 1590
 QY 388 -----ValIleGluAlaProGlyAl 394
 DB 1591 CATCTGCGCTTCACACTCTTGCACCTAGCTAGCTGCAAGTGTATAGAGGCGCTGGTGC 1650
 QY 394 aserLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTrpCys 414
 DB 1651 CTCAAAATGATTGTGGCCCGGCTGTCTCACTGGCGGCGACCGGCTCGGTGTACTGCAG 1710
 QY 414 rAsnAspCysIleLeuLysHisAlaAlaLarMetLysPheLeuSerSerGlyLysG 434
 DB 1711 TAATGACTGATCTCTAACAACGCGCGAGCGACATGAGTCTTCTAAGCTCAGTAAGA 1770
 QY 434 uGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSerLeuPro 454
 DB 1771 ACAGAGCCCAAGCCTAAGAAAGATGAAGATGAAGCCAGAGAGAGAGAGAGAGAGAGAG 1830
 QY 454 sCysGlyAlaGln----- 458
 DB 1831 ATGCGGTGCTCAGCGCTGCTCCAGTGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1890
 QY 459 -AlaGlyIleLysLysSerSerValHisLysArgProAlaProGluLysLysLysLys 478
 DB 1891 ATCAGGCTTGCAGGTG---ACTCTGTGCGGAGCGCGCTTCAAGGAGGAGGAGGAGG 1941
 QY 478 rValLysLysAlaValValValProAlaArgSerGluAlaLeuGlyLysGluAlaAla 497
 DB 1942 GTTGAGATTCTCAGTTTG-----CACGGGGAACCTGCGGCGGCGGCGGCGGCGG 1992
 QY 498 ----CysGluSerSerThrProSerTrpAlaSerAspHisAsnTrpAsnAlaValLys 516
 DB 1993 CCGAATACGCAAGCTCAACAAAGTACTGGGAGGCGCTGCCACTGTATGCTTTGAGGCC 2052
 QY 516 o-----GluLysThrAlaAlaProSerProSerLeuLeuTrpLysCysMetTy 532
 DB 2053 CTGCGCTGCTCCCTGGCTGGCCACGAGGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2095
 QY 532 rHisLeuGlyValGlyLeuLeuAspProSerArgSerPhe----- 545
 DB 2096 -----GTAGGGG---CTGGATCCCTGCGCGCGGATGCGGTGACCTTCAACAGAGGCC 2143
 QY 546 -----TrpIleAlaIleProTrp-----AlaCysProGlyLeu---GlyValAlaAl 560
 DB 2144 TGTCCCTGGGCGCTGTCTCTCTCTGGGAGCGCGCTGTCTCTCTCTCTCTCTCTCTCTCT 2203

QY 1 MetAspLysGlyAspProSerAsnGluAlaProLysAlaIleLysProThrSer 20
DB 232 ATGATGATATAAGGCGACCTGAGCATGAGGAGCACCACCAAGGCTATCAAAACCCACAGT 291
QY 21 LysGluPheArgLysThrTyrPheArgThrThrIleAlaLysArgGluGlyAla 40
DB 292 AAGGAGTTTCAAGAAACCTGGGGTTTTCGAAAGAACCCAGGATTCGCAACGTCGAGGGTGA 351
QY 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
DB 352 GAGACACGAGGCGGACCCAGTCAGTCAGCAACCA-----CAGCAGCAATACCTCTCC 405
QY 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrIle 80
DB 406 CTGCGCCGAGTGGACGCGCAACCAACGTTACTGAGAGCGGTAGAAAGAGTTCTTTACACAG 465
QY 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
DB 466 GTTCGCGCGGAGGAAAGAAATGTCCGGTGTCCCTGGAGATTCAGTCAGTCAGGCGCACA 525
QY 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
DB 526 TCTTCCACAGTCACCTGATGTGGAGACAGCTTCGAGGCGGAGCGTTGAAAGACAGTTCGAG 585
QY 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
DB 586 ATCAGAGTGGCCCTGTATCTGACTCTCTAGGG-----AAAGAACATCTCTCCCTCTTCTGA 642
QY 141 LysValLysGlyLysAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
DB 643 AAGCCAAAGAGGTGAAGAGGAGAGACACCTCTGACAGTCAGTCAGTCAGTCAGTCAGT 702
QY 161 LeuLysGluLeuGlnAsnArgLeuArgLysArgGluGlnGluProThrGluArgPro 180
DB 703 TTGAAGGAACCTTCAGAACCGCTTCGAGAAACGAGAGCAAGAACCTCTGAGAGAGTCC 762
QY 181 LeuLysGlyLysSerArgLeuArgLysArgGluGluGluGluGluGluGluGluGlu 200
DB 763 CTGAGGCGAGTCAGATTCCTGAGGAGAGAGCGCAGAGAGAGAGATTCGCGGAACT 822
QY 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuPro---SerLysGlnGluPro 219
DB 823 ---GGAGTGTCCAAATAGGCTGCGAGCAGGACAGACCTCTCTGTAGAGGAGCGCT 879
QY 220 GluAsnAspGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGlu 239
DB 880 GAGGCTAGTCAGGAGCAGTGTCCAGTCAGAGCAGAGTCAGTATAGTAATCAGTTGGAA 939
QY 240 GlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysPro 259
DB 940 GGGAGGCGACTCAGGGAATACAGAGGAAACCCCGAGGAGCGGCAACCCAAAGCCT 999
QY 260 GluCysGluGlyTyrAspProAlaLeuTyrCysIleCysArgGlnProHisAsnAsn 279
DB 1000 GAGTGTGAGGTTTACGACCCCAATGCTGTACTGATCTGCGCCGAGCCTCAACAAC 1059
QY 280 ArgPheMetIleCysCysAspArgCysGluGluTyrPheHisGlyAspCysValGlyIle 299
DB 1060 AGTTTATGATCTGCTGTATCGGTGTGAGAGTGTTCATGTCGTCGTCGTCGTCGTCGTC 1119
QY 300 SerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCys 319
DB 1120 TCTGAGCGCGAGCGGCTCTCTGGAAGGAACCGGGAAGAGTACATCTGCGCAATATTC 1179
QY 320 ThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLys 339
DB 1180 ACCATTTTGCATGTCAGTGCAGATGAGACAAACCGTAGCGCCACCAATGAGCAGGACTCTGGG 1239
QY 340 TrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLys 359
DB 1240 TGCAGATCTCTGGGTCTGTATGTCACAGACTGACAGCAGCAGTAGGGAACAGTAGAGCAGAG 1299
QY 360 SerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 379

QY 560 aleu 561
DB 2204 TCTG 2207
RESULT 10
AAZ90579 standard; cDNA; 2867 BP.
ID AAZ90579;
AC AAZ90579;
XX
DT 19-JUN-2000 (first entry)
DE Murine death inducer-obliterator 1 (DIO-1) polypeptide encoding cDNA.
XX
KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; murine;
KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW antirheumatic; antiinflammatory; antiproliferative; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 232..2076
FT /tag= a
FT /product= "DIO-1"
XX
PN WO200015787-A1.
XX
XX 23-MAR-2000.
XX
XX 10-SEP-1999; 99WO-GB003019.
XX
XX 10-SEP-1998; 98SE-00003069.
XX 17-SEP-1998; 98US-0100873P.
XX
XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
XX (BANN/) BANNERMAN D G.
XX
XX Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
XX
XX WPI; 2000-271426/23.
XX P-PSDB; AAY67580.
XX
XX New DNA encoding human and murine death inducer-obliterator 1
XX polypeptides, useful in the treatment of cancer, autoimmune diseases,
XX diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
XX hyperproliferative skin disorders.
XX
XX Claim 3; Fig 1B; 27pp; English.
XX
XX The invention provides nucleic acids encoding the human and murine death
XX inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
XX expressed by standard recombinant methodology. The DIO-1 polypeptides,
XX agonists and antagonists are used as a medicament for treating diseases
XX characterized by an alteration in cell death or by hyperproliferation,
XX e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
XX tumours, malignant tumours or hyperproliferative skin disorders. They are
XX also useful in the treatment of metabolic, proliferative or inflammatory
XX conditions. The present sequence represents a cDNA encoding the murine
XX DIO-1 polypeptide
XX
XX Sequence 2867 BP; 763 A; 666 C; 774 G; 664 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1,75e-122 Length: 2867
XX Score: 2112.50 Matches: 422
XX Percent Similarity: 81.40% Conservative: 33
XX Best Local Similarity: 75.49% Mismatches: 94
XX Query Match: 70.68% Indels: 11
XX DB: 3 Gaps: 7
XX
XX US-09-787-016a-3 (1-562) x AAZ90579 (1-2867)

Db 1300 TCCGAGAGACCGGATTAAGGGTAGGATGAGAGGCGCAACCCAGCGGCAAG 1359
Qy 380 LysLysLeuLysIlePheGlnProValIleGluAlaProGlyValAspLysCysIleGly 399
Db 1360 AAAAACTCAAGATATCCAGGCTGTCTGAGAGGCTCTCTGCTCTCTAAATGCAATGCG 1419
Qy 400 ProGlyCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeu 419
Db 1420 CTTGGGTGTCTCAGGTGTAGCACAGCTGACTCTGTGTATGCAAGTAATGCAATGCTC 1479
Qy 420 LysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysPro 439
Db 1480 AACCACGACGACGACCTACCATGATTTCTAAGTTCAAGTAAAGACAAACAAACACCC 1539
Qy 440 LysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAla 459
Db 1540 AAGGAAAAGCTCAACAGCAAGCCAGAAAAGTTCAAGTCTTCCAAAATGCAAGTTCAGGTG 1599
Qy 460 GlyIleLysIleSerSerValHisLysArgProAlaProGluLysIleGluThrVal 479
Db 1600 GGGATTAATCTCTCTGTGCACAGACCTAGCTCTGAGAAAGGGAACCAACCACTG 1659
Qy 480 LysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGlu 499
Db 1660 AAGAAA---GTGATGCTGCTTCCAGAGTGAGACTTCTGGAGGAGGAGGAGCTGTGAG 1716
Qy 500 SerSerThrProSerTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThr 519
Db 1717 AGCAGCACACCATCTCTGGGCAAGTACCACTCAATGCTGTGAAGCCAGAGAGCCA 1776
Qy 520 AlaAlaPro-----SerProSerLeuLeuTyLysCysMetTyHisLeuGlyVal 536
Db 1777 GAGAGCCCACTGCACTCTGCGCCACCCCTATTGAGTAAATGATACGTATCAACCAAGGCT 1836
Qy 537 GlyLeuLeuAspProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGly 555
Db 1837 GGTCTCCAGGCGCCCTCCCATC-ATCTGGGTGGCT---GCCTGGGGCTGTCTAGGA 1888
RESULT 11
AAC77933
ID AAC77933 standard; cDNA; 1764 BP.
XX
AC AAC77933;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:327.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropenic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
OS Homo sapiens.
XX
PN W020005350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US005882.
XX
XX 12-MAR-1999; 99US-0124270P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.
DR P-PSDB; ABA43724.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer.
XX
PS Claim 1; Page 880-881; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC ABA43998 to ABA44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antiaesthetic; antirheumatic; antithyroid; antiviral;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC neutropenic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions; graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and ABA44240 represent sequences used in the exemplification of
CC the present invention
XX SQ Sequence 1764 BP; 468 A; 424 C; 415 G; 453 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 2,22e-97 Length: 1764
Score: 1706.00 Matches: 317
Percent Similarity: 99.37% Conservative: 0
Best Local Similarity: 99.37% Mismatches: 1
Query Match: 57.08% Indels: 2
DB: 3 Gaps: 0

US-09-787-016A-3 (1-562) x AAC77933 (1-1764)

Qy 245 AspileLysAspGluGluProGlyAsp-LeuGlyArgProLysProGluCysGluGlyTy 264
Db 2 GACATCAAGATGAGAGGCTTGGGCGAGCGAGCTTGAATGAGGGTTA 61
Qy 264 rAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArgPheMetileCy 284
Db 62 CGACCCCAAGCGCCCTGATTGCTTGGCGGAGCTCACACACACAGGTTTATGATTG 121
Qy 284 sCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgG 304
Db 122 CTGTGACCGCTGTGAAGATGTTTCATGGCGATTGTGGGCATTTCTGAGGCTCGAGG 181
Qy 304 YArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThrIleLeuGlnVa 324
Db 182 GAGGCTTTTGAAGAAGATGGGAGAGCTATATCTGCCCAACTGCACCATTTCTCAAGT 241
Qy 324 IGLnAspGluThrHisSerGluThrAlaAspGlnGluAlaLysTrpAcgProGlyAs 344
Db 242 GCAGGATGAGACTCATTCAGAAACGGCAGATCAGCAGAGGCTAAATGAGACCTCGAGA 301
Qy 344 pAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspG 364
Db 302 TGCTGATGACCGGATTTGATAGATATAGGAACATAGGAGAGAGTCTAGCGAAGACCA 361
Qy 364 nGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyIleLysLysLeuLysI 384
Db 362 AGGATAAAGGCTAGAAATGAGAAAGCTGCAATCAAGTGGCAGAGAGAACTCAAGT 421
Qy 384 ePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysH 404

Db 422 CTTCCAGCCTGTGATAGAGCGGCTGTGCTCAAAATGATTTGGCCCCGGGTGTGTCA 481
 Qy 404 sValAlaGlnProAspSerValTyrcysSerAsnAspCysileuLysHisAlaAla 424
 Db 482 CGTGGGGCA-CCGAGCTCGTGTACTGCACTATGACTGTATCTCTCAACACGCGCAGC 540
 Qy 424 aThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLys 444
 Db 541 GACAATGAAGTTTCTAGCTCAGGTAAAGAACACAGAACCCAAAGCCTAAAGAAAGATGA 600
 Qy 444 sMetLysProGluLysProSerLeuProLysCysGlyValadlnAlaGlyileLysileSe 464
 Db 601 GATGAGCCAGAGAGGCCAGTTCTCCGAATGCGGTGCTCAGCAGGTATTTAAATCTC 660
 Qy 464 rSerValHisLysArgProAlaProGluLysLysGluThrThrValLysLysAlaValVa 484
 Db 661 TTCTGTGACACAGACAGCTCCAGAAAAAAGACACACACATGCAAGAGCAGTGGT 720
 Qy 484 lValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCysGluSerSerThrProSe 504
 Db 721 GGTCCCTGTGCGGAGTGAAGCATCTCGGAAGGAGCAGCTTGTGAGAGCAGCAGCGTC 780
 Qy 504 rTPAlaSerAspHisAsnTyraAlaValLysProGluLysThrAlaAlaProSerPr 524
 Db 781 GTGGGGAGCGATCACAAATTACATGCAATGAGTAAAGCCAGAAAGACTGTCTCCCTCGCC 840
 Qy 524 oSerLeuLeuTyrcysMetTyrcysMetTyrcysMetTyrcysMetTyrcysMetTyrcys 544
 Db 841 GTCACTGTCTGATAAATGTATGATATCATCCAGGGGTGGCTCTGTGACCCCTCCCGGTC 900
 Qy 544 rPheTTPileAlaIleProTTPAlaCysProGlyLysGlyValAlaAlaLeuLys 562
 Db 901 TTCTGATAGCCATCCCTGGGCTGTCCAGACTGGAGATTGCGACTTTGTGT 955

 RESULT 12
 ID AAS86575 standard; cDNA; 2986 BP.
 AC AAS86575;
 XX 13-FEB-2002 (first entry)
 DT DNA encoding novel human diagnostic protein #22379.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; es.
 XX Homo sapiens.
 OS WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG22388.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 22379; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS54197-AAS94584 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2986 BP; 798 A; 796 C; 781 G; 611 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,11e-82 Length: 2986
 Score: 1467.00 Matches: 294
 Percent Similarity: 86.73% Conservative: 0
 Best Local Similarity: 86.73% Mismatches: 1
 Query Match: 49.08% Indels: 44
 DB: 5 Gaps: 1

 US-09-787-016A-3 (1-562) x AAS86575 (1-2986)
 Qy 235 GluSerLysLeuGluGluLysAlaAlaGlnAspLysLeuAspGluProGlyAspLeu 254
 Db 3 GAGATGAAGTTGGAGGAAAGCGCGCTCAGACATCAAGATGAGGAGCGCTGAGACTTG 62
 Qy 255 GlyArgProLysProGluCysGluGlyTyrcysProAsnAlaLeuTyrcysileCysArg 274
 Db 63 GCGCGACCGAGCGCTCAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCAATTTGCCG 122
 Qy 275 GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluThrPheHisGly 294
 Db 123 CAGCCTCACACACACAGGTTTATGATTTCTGTGACCGCTGTGAGAAAGTTTCATGCGC 182
 Qy 295 AspCysValGlyileSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyrc 314
 Db 183 GATTGTGTGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGCGGAGACTAT 242
 Qy 315 IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp 334
 Db 243 ATCTGCCCAACTGCACCATTTCTGCAAGTGCAGGATGAGACTCATTTCAAGAACGGCAGAT 302
 Qy 335 GlnGlnGluAlaLysTTPArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly 354
 Db 303 CACGAGAGCTAAATGAGAGACTGGAGATGCTGATGCGCACCGATTGTACAGTATAGGA 362
 Qy 355 ThrIleGluGlnLysSerSerGluAspGlnGlyileLysGlyArgIleGluLysAlaAla 374
 Db 363 ACAATAGACAGCAAGTCTAGCGAAGCACAAGGATTAAGGGTAGAATTGAGAAAGCTGCA 422
 Qy 375 AsnProSerGlyLysLysLysLeuLysilePheGlnPro----- 387
 Db 423 AATCCAGTGGCGAGAGAGAACTCAAGATCTTCCAGCTTGGTCCCGGGCTGTGCCACC 482
 Qy 387 ----- 387
 Db 483 CAGCTGCTGTCTCTGCGCAGGATTTGGAAATGCTGTCTAGAAAGCATCTCTGCTTTC 542
 Qy 388 -----ValIleGluAlaProGlyAlaSerLysCysile 398
 Db 543 ACACTCTTGACTGCATTAGCTGCAAGGATAGAGGGCGCTGTGGCTCAAAATGTATT 602

PR	14-AUG-2000;	2000US-0225266P
PR	14-AUG-2000;	2000US-0225270P
PR	14-AUG-2000;	2000US-0225274P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225275P
PR	14-AUG-2000;	2000US-0225275P
PR	18-AUG-2000;	2000US-0226279P
PR	22-AUG-2000;	2000US-0226681P
PR	22-AUG-2000;	2000US-0226682P
PR	22-AUG-2000;	2000US-0226682P
PR	23-AUG-2000;	2000US-0227009P
PR	30-AUG-2000;	2000US-0228294P
PR	01-SEP-2000;	2000US-0229287P
PR	01-SEP-2000;	2000US-0229343P
PR	01-SEP-2000;	2000US-0229344P
PR	01-SEP-2000;	2000US-0229345P
PR	05-SEP-2000;	2000US-0229503P
PR	06-SEP-2000;	2000US-0229513P
PR	06-SEP-2000;	2000US-0230437P
PR	06-SEP-2000;	2000US-0230438P
PR	08-SEP-2000;	2000US-0231242P
PR	08-SEP-2000;	2000US-0231243P
PR	08-SEP-2000;	2000US-0231244P
PR	08-SEP-2000;	2000US-0231413P
PR	08-SEP-2000;	2000US-0231414P
PR	08-SEP-2000;	2000US-0232080P
PR	14-SEP-2000;	2000US-0232393P
PR	14-SEP-2000;	2000US-0232398P
PR	14-SEP-2000;	2000US-0232401P
PR	14-SEP-2000;	2000US-0233063P
PR	14-SEP-2000;	2000US-0233064P
PR	14-SEP-2000;	2000US-0233065P
PR	14-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-0234227P
PR	21-SEP-2000;	2000US-0234228P
PR	25-SEP-2000;	2000US-0234597P
PR	25-SEP-2000;	2000US-0234598P
PR	26-SEP-2000;	2000US-0234584P
PR	27-SEP-2000;	2000US-0235834P
PR	27-SEP-2000;	2000US-0235836P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236368P
PR	29-SEP-2000;	2000US-0236369P
PR	29-SEP-2000;	2000US-0236370P
PR	29-SEP-2000;	2000US-0236370P
PR	02-OCT-2000;	2000US-0236502P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	02-OCT-2000;	2000US-0237039P
PR	13-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239933P
PR	13-OCT-2000;	2000US-0239937P
PR	20-OCT-2000;	2000US-0240960P
PR	20-OCT-2000;	2000US-0241221P
PR	20-OCT-2000;	2000US-0241785P
PR	20-OCT-2000;	2000US-0241786P
PR	20-OCT-2000;	2000US-0241787P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241809P
PR	20-OCT-2000;	2000US-0241809P
PR	01-NOV-2000;	2000US-0244617P
PR	01-NOV-2000;	2000US-0244617P
PR	08-NOV-2000;	2000US-0246474P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0245523P
PR	08-NOV-2000;	2000US-0245524P
PR	08-NOV-2000;	2000US-0245525P
PR	08-NOV-2000;	2000US-0245526P

AAK67051

XX

XX

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

KW cytostatic; gene therapy;

US
XX
HOMER SAFTS .

XX
FN 7000704
7000704-72.

2000

10

PR 04-FEB-2000; 2000US-0180622

PR 02-MAR-2000; 2000US-0186355

PR 17-MAR-2000; 2000US-019007

PR 19=HAF=2000; 2000US=020335
PR 07=HIN=2000; 2000US=020946
PR

28-JUN-2000: 2000US-021513
30-JUN-2000: 2000US-021513

07-JUL-2000: 2000US-021688
PR 07-JUL-2000: 2000US-021688

PR 11-JUL-2000; 2000US-021749

PR 26-JUL-2000; 2000US-02209E

PR 14-AUG-2000; 2000US-022451

14-AUG-2000; 2000US=022521

FR I#-AUG=2000; Z000US-VZ33Z
BR 14-AUG-2000: 2000HS-0225Z

PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG22387.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 22378; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pot_sequences
 XX Sequence 503 BP; 165 A; 107 C; 138 G; 93 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 1-29e-47 Length: 503
 Score: 898.00 Matches: 167
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 30.04% Indels: 0
 DB: 5 Gaps: 0

US-09-787-016A-3 (1-562) x AAS86574 (1-503)

QY 304 GlyArgLeuLeuLysGlnGlyGluAspThrLeuGlyCysProAsnCysThrIleLeuGln 323
 DB 1 GGAGAGGCTTTGGAAAGGAATGGGAGAGACTATATCTGCCCAACTGCACCACTTCGAA 60
 QY 324 ValClnAspGlnThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrpArgProGly 343
 DB 61 GTCCAGGATCAGACTCATTCCAGAAACGGCAGATCAGCAGAGAGCTAAATGGAGACTCGA 120
 QY 344 AsnAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlySerSerGluAsp 363
 DB 121 GATGCTGATGGCCACCATTTGTACAGATAGAACCAATAGACAGAGCTTAGCGAGAC 180
 QY 364 GlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLys 383
 DB 181 CAAGGATTAAGGGGTAGATTGAGAAAGCTGCATATCCAAAGTGGCAGAGAACTCAAG 240
 QY 384 IlePheClnProValIleGluAlaProGlyAlaSerIleCysValIleGlyProClyCysCys 403

Db 241 ATCTTCAGGCCTGTGATAGAGCGCGCTGGTGGCTCAAAATGTAATGGCCCCGGTGCTGT 300
 QY 404 HisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAlaLa 423
 Db 301 CACGTGGCGCAGCCCGACTCGGTGTAAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 360
 QY 424 AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMet 443
 Db 361 GCGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGAGCCCAAGCTTAAGAAAGATG 420
 QY 444 LysMetLysProGluLysProSerLeuProLysCysGlyValAlaGlnAlaGlyIleLysIle 463
 Db 421 AAGATGAAGCCAGAGAGCCCGACTCTTCGAAATGCGGTGCTCAGCGAGTATTAAATC 480
 QY 464 SerSerValHisLysArgPro 470
 Db 481 TCTTCTGTGTCACCAAGAGACCA 501
 RESULT 15
 ACH31915
 ID ACH31915 standard; cDNA; 478 BP.
 XX XX ACH31915;
 AC ACH31915;
 XX 13-OCT-2003 (first entry)
 DT Human endothelial cell cDNA #48.
 DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 DE genome mapping; biodiversity; genetic disorder.
 KW Homo sapiens.
 XX OS
 XX US2003073623-A1.
 XX 17-APR-2003.
 XX 30-JUL-2001; 2001US-00918995.
 XX 30-JUL-2001; 2001US-00918995.
 XX (DRMA/) DRMANAC R T.
 XX (LABA/) LABAT I.
 XX (STAC/) STACHE-CRAIN B.
 XX (DICK/) DICKSON M C.
 XX (JONE/) JONES L W.
 Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
 WPI; 2003-615964/58.
 New polynucleotide sequences obtained from various cDNA libraries, useful
 as hybridization probes, as oligomers for PCR, for chromosome and gene
 mapping, in the recombinant production of protein, or in generating
 antisense DNA or RNA.
 Claim 1; SEQ ID NO 19127; 44pp; English.
 The invention relates to an isolated polynucleotide comprising any one of
 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
 determined by the technique of SBH (sequencing by hybridisation). Also
 included is a purified polypeptide comprising a sequence corresponding to
 a reading frame of the novel polynucleotide. The nucleic acid sequences
 are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was

CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 478 BP; 131 A; 98 C; 150 G; 97 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 5,71e-42 Length: 478
Score: 807.00 Matches: 144
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.00% Indels: 0
DB: 8 Gaps: 0

US-09-787-016A-3 (1-562) x ACH31915 (1-478)

QY	215	SerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAspArg	234
DB	46	AGTAACAGAGAGCCCGAGAACGATCAGCGGGTGTCTCCAGGCTGGGAAGATGACACA	105
QY	235	GluSerLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeu	254
DB	106	GAGAGTAAGTTGGAGGGAAGGGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTG	165
QY	255	GlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArg	274
DB	166	GGCCGACCCGAAGCCTGAATGTGAGGGTTACGACCCCAAGCCCTGTATTGCATTTGCCGC	225
QY	275	GlnProHisAsnArgPheMetIleCysCysAspArgCysGluGluTrpPheHisGly	294
DB	226	CAGCCTCACACAAACAGGTTTATGATTTCCTGTGACCGCTGTGAGAAATGTTTCATGGC	285
QY	295	AspCysValGlyTleSerGluAlaArgGlyArgLeuLeuGluArgGluAspTyr	314
DB	286	GATTGTGTGGGCATTTCAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGGAAGACTAT	345
QY	315	IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp	334
DB	346	ATCTGCCCAACTGCACCATTCCTGCAAGTCACGATGAGACTCATTTCAGAAACGGCAGAT	405
QY	335	GlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly	354
DB	406	CAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCGGATTGTACAGTATAGGA	465
QY	355	ThrIleGluGln	358
DB	466	ACAATGAGCGAG	477

Search completed: May 1, 2004, 17:45:21
Job time : 611 secs

RESULT 2

US-09-764-864-1258
Sequence 1258, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1258
LENGTH: 268
TYPES: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (162)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (209)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (264)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (265)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (268)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1258

Query Match 6.1%; Score 181.5; DB 9; Length 268;

Best Local Similarity 27.7%; Pred. No. 7.1e-05; Mismatches 52; Conservative 29; Gaps 10;

QY	224	GVVSO-AGKDDRESKLEKGAQDIKDEEPDGLRKPKECEGYDPAALYICICQPHNRP	282
DB	29	GVVREKAGSGMEG-----DGSDEPPDAGDS-KSENGENAPIYICICKPDIKCFM	79
QY	283	ICCDCEWFHGDVCVISEARGLLERNGEDYICPNC-----TLQVQ-----	326
DB	80	ICGDNQNFHGDICIRITERMAKAI-----REWCRCREKDKLEIRYHPHKSRERDQNE	135
QY	327	ETHSETADQQAQRWP-GDAGTDTCTSIGTIEQKSSDQIGRIEK-AANPSGKKLKI	384
DB	136	RDSSEPDGSGGRKRPVDPD-----LQRAKSGTGVGNLARGSAPHKSSP---	183
QY	385	FOVIEAP 392	
DB	184	-QLVATP 190	

RESULT 3

US-09-823-187-88
Sequence 88, Application US/09823187
Publication No. US20030096952A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumder, Kundu
APPLICANT: Padigaru, Muralidhar
APPLICANT: Patturajan, Meera
APPLICANT: Shimkets, Richard A
APPLICANT: Spaderna, Steven K
APPLICANT: Spyttek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 88
LENGTH: 669
TYPE: PRT
ORGANISM: Mus musculus
US-09-823-187-88

Query Match 5.8%; Score 174.5; DB 10; Length 669;

Best Local Similarity 20.9%; Pred. No. 0.00072; Mismatches 128; Conservative 77; Gaps 29;

QY	2	DDKDGPSNEZAPKAKPTSKFRKTWGFRTTIAREGAGDAEADPLEPPPPQOOLGSL	61
DB	197	EDSESPSESE-----KTSQDF-----TPEKXTAA-----RPPRGFLG---	220
QY	62	RRSGRPQRTSERVQFLTIARRRRRMPVLSGHTSCPTADTAETASGVSASSET	121
DB	221	---CRKKKVPASDSKADSGAKBFPV---TAQPS---PSSSSSSSSSSSDVSV	272
QY	122	RSQPQASATVK-----ERPASSEKVGDDHDDTSDSDSLTKELQN---	166
DB	273	KKPRGRKPAKPPPKGRPKRPPST-----SSSDSDSGEVDRISEWKR	322
QY	167	---RURRKRQEPTEPLKIGSRURKKRRBEGPAPTVGSEASDTVEGVLPKQ	217
DB	323	RDEBRRLRLRRRREQEELRLRLR-QEREKERKE-RAERGS-----SGE	369
QY	218	EPENDQGVVSGAGKDDRESKLEGKAAQDIKDEEP-GDLGR-----PKPECEGYDPNLY	270
DB	370	LEBDEEPV-----KKSRKARGRTSSSDSEPEGELGKGGKLAKSQLPSSS	419
QY	271	CTCRQPHNNRFMI CDRCEWFHGDVGVISARGLLERNGEDYICPNTILQVQDETHS	330
DB	420	--ARKP-----GQKEKGR-----	431
QY	331	ETADQQAQRWPGDAGTDTCTSIGTIEQKSSDQIGRIEKAANPSGKKLK---	384
DB	432	---PDEKFRARPVKVET-----KRGSEGLSKRGKKEKPSVEERLQKLHSLIK	479
QY	385	FOVIEAFGASKICGP-----QCCHVA-----QPDV-----YCSNDCILKHA	423
DB	480	FALKVDNPDVVKLSALEELGTQVTSQILQKNTDVVATLKIRRYKANKDVMAKAEVY	539
QY	424	-----ATMKFLSSGKQKPKPKMKMKPKPSLPKCGAGAGIKISSV---	467
DB	540	TLKSRVLGPKVYALQKVKAGAKERADNKLREQ-----GEQAPRLADEPST	591
QY	468	KRPAPEKKEETVKKAVVVPARSEALGKEAACESSTPSSWASH-----NYNAVKP---	516

Db 592 DRSAFVNGEATSKGEMNEDRAQEDQDS--EDGPRGSGSSEELHDSPRDNDSPAKPGNER 649

Qy 517 ---EKTAAPSPS 525

Db 650 QDHERTRLASES 661

RESULT 4

US-09-863-776-53

Sequence 53, Application US/09863776

Publication No. US20030198953A1

GENERAL INFORMATION:

APPLICANT: Spytek, Kimberly A

APPLICANT: Majumder, Rumud

APPLICANT: Tchernev, Velizar T

APPLICANT: Mishra, Vishnu

APPLICANT: Padigaru, Muralidhara

APPLICANT: Spaderna, Steven K

APPLICANT: Shenoy, Suresh G

APPLICANT: Rastelli, Luca

APPLICANT: Li, Li

APPLICANT: Taupier, Raymond J

APPLICANT: Gangolli, Esha

TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-020

CURRENT APPLICATION NUMBER: US/09/863,776

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 09/540,763

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/206,679

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,688

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,829

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/207,748

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/207,798

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/208,263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 60/208,831

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/209,451

PRIOR FILING DATE: 2000-06-05

PRIOR APPLICATION NUMBER: 60/210,060

PRIOR FILING DATE: 2000-06-07

PRIOR APPLICATION NUMBER: 60/219,507

PRIOR FILING DATE: 2000-07-20

PRIOR APPLICATION NUMBER: 60/221,337

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/221,927

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: 60/263,135

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: 60/263,688

PRIOR FILING DATE: 2001-01-24

PRIOR APPLICATION NUMBER: 60/263,694

PRIOR FILING DATE: 2001-01-24

NUMBER OF SEQ ID NOS: 155

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 53

LENGTH: 669

TYPE: PRT

ORGANISM: Mus musculus

US-09-863-776-53

Query March 5.8% Score 174.5; DB 10; Length 669;

Best Local Similarity 20.9%; Pred. No. 0.00072;

Matches 128; Conservative 77; Mismatches 182; Indels 225; Gaps 29;

Qy 2 DDKGDPSEBAPKAIKTSKFRKTWGRFTTIARREGAGDAADPLEPPPPQQQIGLIL 61

Db 187 EDSPSPSESE-----KTSDDQDF-----TPKKTAA-----RPPRRGFLG--- 220

Qy 62 RRSGRQPKTERVROFLTIARRRRRRMPVLEDSGETSCPATDAETASGSEVESASET 121

Db 221 ---GRKKKKVPSASDSKADSDGAKSEFVV---TAQPS---PSSSSSSSSSSSDSVS 272

Qy 122 RSGPOSASTAVK-----ERPASSEKVKGGDDHDDTSDSDSDGLTILKEIQN--- 166

Db 273 KCPFRGRAPAEKPPKPRGRAPKPERPEST-----SSSDSDSDSGEVDRISEWKR 322

Qy 167 -----RLRRKREOQPTERPLKGIOSRLRKRRREGPARTEVGSSEASDTVEGVLPKQ 217

Db 323 RDEERRRELEARRRREOQEELRLAE-QERBEKERKE-RAERGGG-----SGE 369

Qy 218 EPENDQGVVSQAGKDDRESKLEGGAAQDIKDEEP-GDILGR-----PKPECEGYDPNLY 270

Db 370 ELEDEEFV-----KCRSRKARGRTTPSSSDSEPEHGLKSGKGLAKKSQLPSSS- 419

Qy 271 CICROPHNNRPMICCDRCBFWFHGDCVIGISEARGRLRLRNGSDYICPNCITILQVODETHS 330

Db 420 ---ARKP-----GQKEKRGK----- 431

Qy 331 ETADQOEAKWRPGDADGTDCTSIGTIEQKSSSEDQGIKRIEKAANPSGKKLKLK-----I 384

Db 432 ---PPEKPRARPVKVERT-----RKRSEGLSRKRGKKEKPSVVERLQKLHSEIK 479

Qy 385 PQVTEAPGASKCIGP---GCHVA-----QPDVS-----YCSNDCILKHA-- 423

Db 480 FALKVNDPVDVKLSALEBELGLTVTSQILQKNTDVAATLKIRRYKANKDVMAKAAEVY 539

Qy 424 -----ATMKFLSSGKEQFPKPKKMKKPKPKPSLPKCAQAGIKSSVH--- 467

Db 540 TRLSRVLPKPVKVEALQKVNKAGAKERADNEKLEQOP-----GEQAPRELAEDPST 591

Qy 468 KRPAPKEKETTVKKAVVVPARSEALGKEAACBSSTPSMASDH-----NYNAVKP- 516

Db 592 DRSAFVNGEATSKGEMNEDRAQEDQDS--EDGPRGSGSSEELHDSPRDNDSPAKPGNER 649

Qy 517 ---EKTAAPSPS 525

Db 650 QDHERTRLASES 661

RESULT 5

US-09-823-187-24

Sequence 24, Application US/09823187

Publication No. US20030096952A1

GENERAL INFORMATION:

APPLICANT: Burgess, Catherine

APPLICANT: Gusev, Vladimir Y

APPLICANT: Liu, Xiaohong

APPLICANT: Majumder, Rumud

APPLICANT: Padigaru, Muralidhar

APPLICANT: Patturajan, Meera

APPLICANT: Shimkets, Richard A

APPLICANT: Spaderna, Steven K

APPLICANT: Spytek, Kimberly

APPLICANT: Taupier, Raymond J

TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 159666-745

CURRENT APPLICATION NUMBER: US/09/823,187

CURRENT FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/193,339

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/193,205

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/195,343

PRIOR FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: 60/195,088

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: 60/195,005

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: 60/195,792

PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 676
TYPE: PRT
ORGANISM: Homo sapiens
US-09-823-187-24

Query Match 5.7%; Score 169; DB 10; Length 676;
Best Local Similarity 20.4%; Pred. No. 0.0018;
Matches 115; Conservative 78; Mismatches 209; Indels 162; Gaps 23;

QY 1 MDDKGDPSNEBAPKAIKPTSKFRKTWGFRRTTIKREGAGDAADPLEPPPPQOQLGLS 60
DB 140 MESDSDSKSDNSGLK-----RKTAL-KVSVSKARKASSDLDOASVSPSEENSES 192
QY 61 LRSGRQPKRTVEQPLTI-ARRG-----RRMPVSLSDSGEPTSC-----PATDAE 108
DB 193 SSESEKTSQDFTPEKKAAPRRGPKGKPAEPLPKPRGKPKPPEPPSS-----SS 302
QY 109 TASGSESVESAS-----ETRSQPSASTAVK-----ERPASSEKVKGGDDHDDT 151
DB 253 SASSSSSSSSSSDSDSVKVPKPRGKPAEPLPKPRGKPKPPEPPSS-----SS 302
QY 152 SDSDSGL-TLKELONR-----LRKKEQETPEPLKIQSLAKKRE---EGPA 198
DB 303 SDSDSDEVDRISEWKRRDEARRLEARRRREBEELRLREQEKEKERRERADRGEA 362
QY 199 E--TVGSEASDTVEGVLPKQEPENDQGVWSQAGKD-DRESKLEGKAAQDINDEBPGDLG 255
DB 363 ERGSGSGSDELRDEDDFVKKGKRGKRGKGGPPSSSDSSEPAELEREAKKSAK----- 414
QY 256 RPKPCBGYDPNALYICICRQPHNNRFTCCDCRCEWFHGDGCVGISEARGRLLENGEYI 315
DB 415 --KQSSSTEP-----ARKPG----- 428
QY 316 CPMCTILQVDETHSETAQOAKWRPGDADGTDCTSIGTIBQKSSDQGIKRIEKAAN 375
DB 429 -----QKERVPEPEKQA--RPVKVET-----RKSEGFSDNRKVEKKKE 468
QY 376 PSKKKKL-----IPQVIEAPGASKGICGOCCHVAOPDSVYCSNDCILKH--AAATWK 427
DB 469 PSVBEKQLKHLSEIKFALKVDSFDVKRCLNA-----LEELGLTQVTSQILQKNTDVTATLK 524
QY 428 FL---SSGKEQKPKPKKMKPEKPSLPKCGAQAQGIKISSVHKRPAPKPKETTAKAVV 484
DB 525 KIRRYKANKDVMEKAAEVTVLRSVLGPKIEAVQKVKAGKEKKEKBEK----- 575
QY 485 VPARGSEALGKAACSSSTPSWASD 508
DB 576 --AGEELAGEELAGEAPQEAED 597

RESULT 6

US-09-863-776-14
Sequence 14, Application US/09863776
Publication No. US20030198953A1
GENERAL INFORMATION:
APPLICANT: Spytex, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Mishra, Vishnu
APPLICANT: Padigaru, Muralidhara
APPLICANT: Spaderna, Steven K

APPLICANT: Shenoy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Li, Li
APPLICANT: Taupier, Raymond J
APPLICANT: Gangolli, Bsha
TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-020
CURRENT APPLICATION NUMBER: US/09/863,776
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 09/540,763
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/206,679
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,688
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/206,829
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,748
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/207,798
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/208,263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 60/208,831
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 60/209,451
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/210,060
PRIOR FILING DATE: 2000-06-07
PRIOR APPLICATION NUMBER: 60/219,507
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/221,337
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/221,927
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 60/263,135
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,688
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,694
PRIOR FILING DATE: 2001-01-24
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 676
TYPE: PRT
ORGANISM: Homo sapiens
US-09-863-776-14

Query Match 5.7%; Score 169; DB 10; Length 676;

Best Local Similarity 20.4%; Pred. No. 0.0018;
Matches 115; Conservative 78; Mismatches 209; Indels 162; Gaps 23;

QY 1 MDDKGDPSNEBAPKAIKPTSKFRKTWGFRRTTIKREGAGDAADPLEPPPPQOQLGLS 60
DB 140 MESDSDSKSDNSGLK-----RKTAL-KVSVSKARKASSDLDOASVSPSEENSES 192
QY 61 LRSGRQPKRTVEQPLTI-ARRG-----RRMPVSLSDSGEPTSC-----PATDAE 108
DB 193 SSESEKTSQDFTPEKKAAPRRGPKGKPAEPLPKPRGKPKPPEPPSS-----SS 302
QY 109 TASGSESVESAS-----ETRSQPSASTAVK-----ERPASSEKVKGGDDHDDT 151
DB 253 SASSSSSSSSSSDSDSVKVPKPRGKPAEPLPKPRGKPKPPEPPSS-----SS 302
QY 152 SDSDSGL-TLKELONR-----LRKKEQETPEPLKIQSLAKKRE---EGPA 198
DB 303 SDSDSDEVDRISEWKRRDEARRLEARRRREBEELRLREQEKEKERRERADRGEA 362
QY 199 E--TVGSEASDTVEGVLPKQEPENDQGVWSQAGKD-DRESKLEGKAAQDINDEBPGDLG 255
DB 363 ERGSGSGSDELRDEDDFVKKGKRGKRGKGGPPSSSDSSEPAELEREAKKSAK----- 414

311 GEDVICPNTILQVDETHSETADQOAKWRPGDADGTDCTISGTHOKSSEDOQIKGRI 370
429 -----QKGRVRPEKQAK--PVKVERT-----RKRSEGFSDMRKV 463
371 EKAANPSGKKLK-----IFQPVIEAPGASKICPGCCHVAQPDVSVYCSNDCLKH--A 422
464 EKKSEPSVEELKXKLSHSEIKFALKVDSFVCRCLNA-----LEELGTQVTSQILQKNTDV 519
423 AATMKFL---SSGKEQKPKPKMKMKPEKPSLPCGAQAQIKLSSVHKRPAPKPKETT 479
520 VATLKIRRYKANDVMEKAAAEVYTRLSRVLGPKIEAVQVKNKAGMEKEKAEKL----- 575
480 KKAUVPARSEALGKEAACSSSTPSWASD 508
576 -----AGEELAGEELAGEAPQEKAD 597
RESULT 8
US-09-823-187-87
Sequence 87, Application US/09823187
Publication No. US2003009652A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine
APPLICANT: Gusev, Vladimir Y
APPLICANT: Liu, Xiaohong
APPLICANT: Majumdar, Kumud
APPLICANT: Padigaru, Muralidhar
APPLICANT: Patturajan, Meera
APPLICANT: Shinkete, Richard A
APPLICANT: Spaderma, Steven K
APPLICANT: Spytek, Kimberly
APPLICANT: Taupier, Raymond J
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-745
CURRENT APPLICATION NUMBER: US/09/823,187
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/193,339
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/193,205
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/195,343
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: 60/195,088
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,005
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/195,792
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: 60/196,556
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: 60/197,081
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/197,525
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/197,087
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 103
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 87
LENGTH: 678
TYPE: PRT
ORGANISM: Mus musculus
US-09-823-187-87
Query Match 5.7%; Score 169; DB 10; Length 678;
Best Local Similarity 20.1%; Pred. No. 0.0016;
Matches 132; Conservative 80; Mismatches 182; Indels 264; Gaps 32;
QY 2 DDKGDPSNEAPKAIKPTSFRKTWGFRTT-----IAKREGAGDAEADPLEPPPPQ 55
DB 143 DSDSDKSSDHS-----GLKRTPLVKSVSVKRARRASSDLDQASVSESE 187

256 RPKPECEGYDNPALYICIQPHNNRPMICCDRCSEWFHDCVGISEARGELLERNGEYI 315
415 ---KPSSSTEP-----AKPG----- 428
316 CPNCTILQVDETHSETADQOAKWRPGDADGTDCTISGTHOKSSEDOQIKGRIEKAAN 375
429 -----QKGRVRPEKQAK--PVKVERT-----RKRSEGFSDMRKVKEKKE 468
376 PSKKKKLK-----IFQPVIEAPGASKICPGCCHVAQPDVSVYCSNDCLKH--AAA 427
469 PSVEELKXKLSHSEIKFALKVDSFVCRCLNA-----LEELGTQVTSQILQKNTDVVATLK 524
428 FL---SSGKEQKPKPKMKMKPEKPSLPCGAQAQIKLSSVHKRPAPKPKETTAKAVV 484
525 KIRRYKANDVMEKAAAEVYTRLSRVLGPKIEAVQVKNKAGMEKEKAEKL----- 575
485 VPARGSEALGKEAACSSSTPSWASD 508
576 ---AGEELAGEELAGEAPQEKAD 597
RESULT 7
US-10-203-708-28
Sequence 28, Application US/10203708
Publication No. US20030149238A1
GENERAL INFORMATION:
APPLICANT: SMITHKLINE BEECHAM CORPORATION
APPLICANT: SMITHKLINE BEECHAM P.L.C.
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP50013
CURRENT APPLICATION NUMBER: US/10/203,708
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: PCT/US01/04703
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,172
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 60/186,084
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 676
TYPE: PRT
ORGANISM: Homo sapiens
US-10-203-708-28
Query Match 5.7%; Score 169; DB 14; Length 676;
Best Local Similarity 20.2%; Pred. No. 0.0016;
Matches 115; Conservative 78; Mismatches 200; Indels 176; Gaps 23;
QY 2 DDKGDPSNEAPKAIKPTSFRKTWGFRTT-----IAKREGAGDAEADPLEPPPPQ 55
DB 143 DSDSDKSSDHS-----GLKRTPLVKSVSVKRARRASSDLDQASVSESE 187
56 QLGLSLRRSGQPKRTVROFLTI-ARRG-----RRSPVLEDSGFTSC-----P 103
188 ENSESSSESEKTSQDPTPEKKAVALRPRGLGRKKKKKAPASDSKADSDGAKPEP 247
104 ATDAETASEGVSAS-----ETESGQSASTAVK-----ERPASSEKVGKD 146
248 VAMARSASSSSSSSSSDSVSVKPKPRGKPAEKPLPKPRGKPKPRPSS----- 300
147 DHDITSDSDGL--TLKELQNR-----LRKRQEPTEPLKIGIQLRKRRE-- 194
301 ---SSSDSDSDVDRISEKWRDREARRELEARRRREQEBELRLRREKEKERRERA 357
195 -EGPAP--TVGSEASDTEVGLPSKQBPENDGVVSOAGKD--DEESKLEKGAQDIDEE 250
358 DRGEAERGSGSGDELEFREDPEPKGRGKRGKRGFPSSDSSEPAELERKAKSAK--- 414
251 PGDLGRPKPECEGYDNPALYICIQPHNNRPMICCDRCSEWFHDCVGISEARGELLERN 310
415 -----KPSSTEP-----AKPG----- 428

```
QY 56 QLGLSLRSGRQPKRTERVE-----QFLTIAR--RGRSRMPVSLDSSEPT--SCPA- 104
D 188 -----DSESESESEKSDQFTPEKTAARPRRG-----PLGGRKKKHTGVACPK 235
QY 105 -----TDAETASGSEV-----SASETSQPSASTAVK----- 133
D 236 VPSASDSDKADSDGAKKEPVVTAQPSPPSSSSSSSSSDSDSVVKKPPGRKPAKPP 295
QY 134 -----ERPASSEKVGDDHDDTSDSDGLTKELQ-----RLRKK 171
D 296 PKPRGRPKPPEPST-----SSSDSDSGEVDRISWKRRDERRERLEARR 345
QY 172 REQEPTEPLKGIQSLRKKRREBEPATVGSASDTVEGLVPSQEPENDQGVSOQAK 231
D 346 REQEEELRLRE-QEREKERRKE-RAERCGS-----SGELEDEBPV----- 386
QY 232 DDRESKLGKAAQDIKDEEP-GDLGR-----PKPCEGYDPNALYICROPHNNRPMIC 284
D 387 KGRSRKARGTSPSSSDSEPEGLKGLAKKSQIPGSES-----ARKP----- 432
QY 285 CDRCEWFHGDVGISEARGLLERNGEDYICPNCITLOVDTHSETADQQAQKVRPGD 344
D 433 -----GQEKRR-----PDEKPRARPVK 451
QY 345 ADGTDCTSIGTIRKSSSEDOGIGRIEKAANPSGKKLK-----IFQPVIEAPGASKCI 398
D 452 VERT-----RKSEGLSLERKKEKKEPSVEERLQKLHSEIKFALKVNDPVRKCL 502
QY 399 GP-----GCCHVA-----QDSV-----YCSNDCILKHA-----A 424
D 503 SALEBELGLQVTSQILQKNTDVVATLKKIRRYKANDVMKAAAEVYTRLSRVLGPKVEA 562
QY 425 TMKFLSSGKQPKPKKKKKKPKTSLPKCGAAGIKISSVH-----KRPAPKKTETVVK 481
D 563 LQKNRAGAKERADNEKLEEQ-----GQAPRELAEDPSTRGAPVNGEATSQK 614
QY 482 AVVVPARSEALGKEACESTPWSADH-----NNAVKP-----EKTAAPS 525
D 615 GENMEDRAQDQGGDS--EDGPGCSSEELHSDPRNSDPAKPNRQHERTRLASES 670

RESULT 9
US-09-863-776-52
; Sequence 52, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderne, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangolli, Edna
; TITLE OF INVENTION: No. US20030198953A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-020
; CURRENT APPLICATION NUMBER: US/09/863, 776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206,679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206,829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207,798
; PRIOR FILING DATE: 2000-05-30
```

```
; PRIOR APPLICATION NUMBER: 60/208,263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208,831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209,451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210,060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219,507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221,337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221,927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263,135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,688
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263,694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-863-776-52
```

```
Query Match 5.7%; Score 169; DB 10; Length 678;
Best Local Similarity 20.1%; Pred. No. 0.00187;
Matches 132; Conservative 80; Mismatches 182; Indels 264; Gaps 32;

QY 2 DDGDDPNEBAPKAIPKTSKFRKTWGRFTT-----IAKREGAGDAEADPLEPPPOQ 55
D 143 DSDSDKSSDHS-----GLKKTPLVKVSVSKRARRASSOLDQASVSPSE 187
QY 56 QLGLSLRSGRQPKRTERVE-----QFLTIAR--RGRSRMPVSLDSSEPT--SCPA- 104
D 188 -----DSESESESEKSDQFTPEKTAARPRRG-----PLGGRKKKHTGVACPK 235
QY 105 -----TDAETASGSEV-----SASETSQPSASTAVK----- 133
D 236 VPSASDSDKADSDGAKKEPVVTAQPSPPSSSSSSSSSDSDSVVKKPPGRKPAKPP 295
QY 134 -----ERPASSEKVGDDHDDTSDSDGLTKELQ-----RLRKK 171
D 296 PKPRGRPKPPEPST-----SSSDSDSGEVDRISWKRRDERRERLEARR 345
QY 172 REQEPTEPLKGIQSLRKKRREBEPATVGSASDTVEGLVPSQEPENDQGVSOQAK 231
D 346 REQEEELRLRE-QEREKERRKE-RAERCGS-----SGELEDEBPV----- 386
QY 232 DDRESKLGKAAQDIKDEEP-GDLGR-----PKPCEGYDPNALYICROPHNNRPMIC 284
D 387 KGRSRKARGTSPSSSDSEPEGLKGLAKKSQIPGSES-----ARKP----- 432
QY 285 CDRCEWFHGDVGISEARGLLERNGEDYICPNCITLOVDTHSETADQQAQKVRPGD 344
D 433 -----GQEKRR-----PDEKPRARPVK 451
QY 345 ADGTDCTSIGTIRKSSSEDOGIGRIEKAANPSGKKLK-----IFQPVIEAPGASKCI 398
D 452 VERT-----RKSEGLSLERKKEKKEPSVEERLQKLHSEIKFALKVNDPVRKCL 502
QY 399 GP-----GCCHVA-----QDSV-----YCSNDCILKHA-----A 424
D 503 SALEBELGLQVTSQILQKNTDVVATLKKIRRYKANDVMKAAAEVYTRLSRVLGPKVEA 562
QY 425 TMKFLSSGKQPKPKKKKKKPKTSLPKCGAAGIKISSVH-----KRPAPKKTETVVK 481
D 563 LQKNRAGAKERADNEKLEEQ-----GQAPRELAEDPSTRGAPVNGEATSQK 614
QY 482 AVVVPARSEALGKEACESTPWSADH-----NNAVKP-----EKTAAPS 525
```

Db 615 GENMEDAQRDQDS--EDGPRGSSSELHSDSPRDNSDPAKPGNVERQDHERTRLASES 670

RESULT 10

US-10-203-708-29

Sequence 29, Application US/10203708

Publication No. US20030149238A1

GENERAL INFORMATION:

APPLICANT: SMITHKLINE BEECHAM CORPORATION

APPLICANT: SMITHKLINE BEECHAM P.L.C.

FILE OF INVENTION: NOVEL COMPOUNDS

FILE REFERENCE: GP50013

CURRENT APPLICATION NUMBER: US/10/203.708

CURRENT FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: PCT/US01/04703

PRIOR FILING DATE: 2001-02-14

PRIOR APPLICATION NUMBER: 60/182,172

PRIOR FILING DATE: 2000-02-14

PRIOR APPLICATION NUMBER: 60/186,084

PRIOR FILING DATE: 2000-02-29

NUMBER OF SEQ ID NOS: 46

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 29

LENGTH: 717

TYPE: PRT

ORGANISM: Homo sapiens

US-10-203-708-29

Query Match 5.7%; Score 169; DB 14; Length 717;

Best Local Similarity 20.2%; Pred. No. 0.0019;

Matches 115; Conservative 78; Mismatches 200; Indels 176; Gaps 23;

Qy 2 DDKGDPNBEAPKAITSKEFTWTFRTT-----IAKRGAGDAEADPLEPPPPQ 55

Db 184 DSDSDKSDNS-----GLKRTPAKMSVSKRAKASDLDQASVSPSE 228

Qy 56 QLGLSLRSGRQPKTERVEQFLTI-ARRRG-----RRSMFVSLSDSGEPTSC--P 103

Db 229 ENSESSSEKTSQDFTPEKAAVAPRPGPLGRRKKKAPASDSDSKADSDGAKPEP 288

Qy 104 ATDAETASBGVSAS-----ETSGQSASTAVK-----ERPASSEKVGGD 146

Db 289 VAMARSASSSSSSSSSDSVSVKPRGRKPAEKPLPKPRKPKPRPPSS----- 341

Qy 147 DHDTSDSDGL-TLXLQNR-----LREKREQEPTEPLKGIQSLRKRRE-- 194

Db 342 ---SSSDSDSDVDRISEWKRDRARRELEARRRQEEELRLRQEKKEKRRERA 398

Qy 195 -EGPAB--FVSGEASDVTVEGLPSQBPENDQGVVSQAGKQ-DRESLEGKAAQDIKDE 250

Db 399 DRGEARGSGSGDELREDDDEPVKGRKGRGRGPPSSSDSEFEAELEAKKSAK--- 455

Qy 251 PGDLGRPKPECEGYDPNALYICIQPHNNRPMICCDRCBWFHGDGVGISEARGLLERN 310

Db 456 -----KPOSSSTEP-----ARKPG----- 469

Qy 311 GEDYICPNTLLQVDETHSETAQQAQKWPBGDADGDTCTSICTIQKSSDQGIKRI 370

Db 470 -----QKGRVPEEKQQAQ--PVKVERT-----RKSESGFSMDRKV 504

Qy 371 EKANPSSKKL-----IPQVIEAPGASKCIQPGCCHVAQSDSVYCSNDILKH--A 422

Db 505 EKKKEPSVEEKQLHSEIKFALKVDSPVVKRCINA-----LEEIGTLQVTSQILQKNTDV 560

Qy 423 AATMKFL---SSGKEQKPKPKKMKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 479

Db 561 VATUKKIRYKANDVMEKAEVTRLSRVLGPKIEAVQKVNAGMEKEKAEKL----- 616

Qy 480 KCAVVVPARSALGKBAACESSTSNASD 508

Db 617 -----AGEELAGEEAPQEKAD 638

RESULT 11

US-09-863-776-16

Sequence 16, Application US/09863776

Publication No. US20030198953A1

GENERAL INFORMATION:

APPLICANT: Spytek, Kimberly A

APPLICANT: Majumder, Kumud

APPLICANT: Tchernev, Velizar T

APPLICANT: Mishra, Vishnu

APPLICANT: Padigaru, Muralidhara

APPLICANT: Spaderna, Steven K

APPLICANT: Shenoy, Suresh K

APPLICANT: Rastelli, Luca

APPLICANT: Li, Li

APPLICANT: Taupier, Raymond J

APPLICANT: Gangolli, Esha

FILE OF INVENTION: No. US20030198953A1

FILE REFERENCE: 21402-020

CURRENT APPLICATION NUMBER: US/09/863,776

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 09/540,763

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/206,679

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,688

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/206,829

PRIOR FILING DATE: 2000-05-24

PRIOR APPLICATION NUMBER: 60/207,748

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/207,798

PRIOR FILING DATE: 2000-05-30

PRIOR APPLICATION NUMBER: 60/208,263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 60/208,831

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/209,451

PRIOR FILING DATE: 2000-06-05

PRIOR APPLICATION NUMBER: 60/210,060

PRIOR FILING DATE: 2000-06-07

PRIOR APPLICATION NUMBER: 60/219,507

PRIOR FILING DATE: 2000-07-20

PRIOR APPLICATION NUMBER: 60/221,337

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/221,927

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: 60/263,135

PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: 60/263,688

PRIOR FILING DATE: 2001-01-24

PRIOR APPLICATION NUMBER: 60/263,694

PRIOR FILING DATE: 2001-01-24

NUMBER OF SEQ ID NOS: 155

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 16

LENGTH: 667

TYPE: PRT

ORGANISM: Homo sapiens

US-09-863-776-16

Query Match 5.6%; Score 167; DB 10; Length 667;

Best Local Similarity 20.5%; Pred. No. 0.0024;

Matches 115; Conservative 78; Mismatches 208; Indels 160; Gaps 23;

Qy 1 MDKGPDSNEAPKAITSKEFTWTFRTTIAKRGAGDAEADPLEPPPPQQLGLS 60

Db 140 MESDSDSDKSDNSGLK-----RKTPLAL-KVSVSKRAKASDLDQASVSPSEENSES 192

Qy 61 LRSGRQPKTERVEQFLTI-ARRRG-----RRSMFVSLSDSGEPTSC-----PATDAE 108

Db 193 SSESRTSQDFTPEKAAVAPRPGPLGRRKKKAPASDSDSKADSDGAKPEPVAMAR 252

143	Db	DSDSKSDNS	-----GLRKT	PALXMSVSKRARKASSDLDQASVSPSEE	187
56	Qy	QLGLSLRRBGRQKPTERVEQFLTI-ARERG	-----RRSMPVSLSDSGEPTSC	-----P	103
188	Db	ENSESSSEKTSQDQFTTEKAAVAPRPRGIGRKKKAPASDSDKADSDGAKPEP	247		
104	Qy	ATDAETVASSGSVESAS	-----ETRSGPQSGASTAVK	-----ERPASSEKVKGGD	146
248	Db	VAMARSASSSSSSSSSDSDSVSKPGRKPAEKLPKPRGRKPKPERPSS	300		
147	Qy	DHDDTSDSDGL-TLKEZLQNR	-----LFRKEQEPTEPLKGIQSLRKREE	-----	194
301	Db	SSSDSDSDVDRISEKMRDEAPRRELEARRRQHEELARLRBQSEKKEKRRERA	357		
195	Qy	BOPAB-TVGBEASVTVGVLPSKQBPNDQGVSVQAGKD-DRSKLEKGAQDIXDBE	250		
358	Db	DRGEARGGSGGDELREDDBPVKRGRKGRGPPSSDSPEAFELERAKKSAK	414		
251	Qy	PGDLGRKPECEGYDPNALYCI-CROPHNKFMICCDRCBEPFHGDCVGISSEARGLLERN	310		
415	Db	KQSSSTEP	-----ARKFG	-----	428
311	Qy	GEDYICPNCITLQVQETHSFTADQOBAKWRPGDADGTCTSGTTEQKSKSEBQIGKRI	370		
429	Db	QBKEVRPEEKQOAK-PVKVERT	-----RKSEGSFMDRKV	463	
371	Qy	EKAANPSGKKLK	-----IFQPVTEAPGASKICIGPCCHVQAPDSVYCSNDGILKH	-A	422
464	Db	EKKQEPSVEBKQLHLHSEIKFALKVDSDPVKRCLNA	-----LEELGITQVTSQILQKXTDV	519	
423	Qy	AATMKEF	-----SSGEQKPKPEKMKWKPKPSLPKCGAQAGIKLSSVHKPEAPKKEITTV	479	
520	Db	VATLKIRRYKANKOMEAEEVYTLKSRVLGPKLIEAVQKVNKAGMBKEKAEKL	575		
480	Qy	KXAVVVVPARSEALGKEAACESSTPFSWASD	508		
576	Db	AGEELAGREAPOEKAEKDPSTD	597		

Search completed: April 30, 2004, 08:57:36
Job time : 52 secs

[illegible]

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:52:47 ; Search time 22 Seconds

(without alignments)

1318.809 Million cell updates/sec

Title: US-09-787-016A-3

Perfect score: 2989

Sequence: 1 MDDKDPSENEAPRAIKPTS.....RSFWIAIPWACPGIGVAALC 562

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/prodata/2/iaa/PCITUS_COMB.pep.*

6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174.5	5.8	2289	3	US-09-051-019-2 Sequence 2, Appli
2	164.5	5.5	1075	4	US-09-252-991A-18387 Sequence 18387, A
3	161.5	5.4	1780	1	US-08-769-309A-5 Sequence 5, Appli
4	161.5	5.4	2138	3	US-08-994-570-5 Sequence 5, Appli
5	161.5	5.4	2138	4	US-09-257-179-80 Sequence 80, Appl
6	158.5	5.3	779	4	US-10-164-595-56 Sequence 56, Appl
7	157	5.3	1404	4	US-08-801-308-1 Sequence 1, Appli
8	152.5	5.1	735	4	US-10-164-595-80 Sequence 80, Appl
9	152.5	5.1	784	4	US-10-164-595-79 Sequence 79, Appl
10	152.5	5.1	843	4	US-10-164-595-54 Sequence 54, Appl
11	152.5	5.1	1596	4	US-08-978-277A-4 Sequence 4, Appli
12	149.5	5.0	2468	4	US-09-976-594-726 Sequence 726, App
13	149.5	5.0	8991	1	US-08-714-741-32 Sequence 32, Appl
14	148.5	5.0	1805	1	US-07-853-913-2 Sequence 4, Appli
15	148.5	4.9	1618	1	US-07-853-913-4 Sequence 54, Appli
16	143.5	4.8	219	2	US-08-557-309B-54 Sequence 27005, A
17	140.5	4.7	1427	4	US-09-252-991A-27005 Sequence 2, Appli
18	138.5	4.6	1061	4	US-09-762-481B-2 Sequence 52, Appl
19	137.5	4.6	442	3	US-08-834-306-52 Sequence 52, Appl
20	137.5	4.6	442	3	US-08-993-674A-52 Sequence 52, Appl
21	137.5	4.6	442	4	US-08-236-376-52 Sequence 32447, A
22	136.5	4.6	700	4	US-09-252-991A-32447 Sequence 6, Appli
23	135.5	4.5	1848	3	US-08-236-791-6 Sequence 6, Appli
24	135.5	4.5	1848	4	US-09-839-996-6 Sequence 6, Appli
25	135.5	4.5	1848	4	US-10-090-505-6 Sequence 6, Appli
26	135.5	4.5	1848	5	PCT-US95-10661A-6 Sequence 2, Appli
27	133.5	4.5	803	3	US-09-083-035-2 Sequence 2, Appli

US-09-051-019-2

Sequence 2, Application US/09051019

Patent No. 6103229

GENERAL INFORMATION:

APPLICANT: KAHMANN, Regine and QUADBECK-SERGER, Claudia

TITLE OF INVENTION: Regulatory gene from Ustilago maydis

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Keil & Weinkauff

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage

COMPUTER: IBM AT-compatible, Pentium processor

OPERATING SYSTEM: Windows 98

SOFTWARE: WordPerfect version 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/051,019

FILING DATE: 31-MAR-1998

CLASSIFICATION:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2289 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-09-051-019-2

Query Match 5.8%; Score 174.5; DB 3; Length 2289;

Best Local Similarity 20.1%; Pred. No. 5.2e-06;

Matches 120; Conservative 85; Mismatches 184; Indels 209; Gaps 28;

QY 28 GFRITTIAXREGAGDADADPLEPPPPQOGLSLRSGRQPKETVEQVPLTIARRGR 87

DB 1387 GSRLLTADNR---DRELD---KLVERVEDATDPADQKP 1420

QY 88 SMPVLSDEGPTSCPATDAETASGVSR---SASTRSGPOS---AS 129

DB 1421 NANCICRSMPIAIPSSSGAECRCRVYHLSCKIVRSEVSRAGGWVCPFCWYGSA 1480

QY 130 TAVKERPASS---EKVKGDDH---DDTSDSDSDGHTLKLQNLRRKREQETEPKLG 183

DB 1481 PFLKREKATSIADLSKLVYDQDHRDQFKFLPLEWDAISEVV--AKAKGFETAAKRMKT 1538

QY 184 IQRLRKRRREE-----GPAS-----TVGSEAST----- 208

DB 1539 L-SLARDDQKVILAWLRRSGICPVDTLGPKEVNMMLISENLLALGSGQGGAAAPV 1597

28	133.5	4.5	1298	2	US-08-590-473-2	Sequence 2, Appli
29	133.5	4.5	1298	3	US-09-259-821A-2	Sequence 2, Appli
30	133.5	4.5	1298	3	US-08-843-659-2	Sequence 2, Appli
31	132.5	4.4	657	3	US-08-893-852A-3	Sequence 3, Appli
32	132.5	4.4	657	3	US-08-821-818-3	Sequence 3, Appli
33	132.5	4.4	657	4	US-09-052-753B-3	Sequence 3, Appli
34	130	4.3	1964	2	US-08-790-912-3	Sequence 2, Appli
35	130	4.3	2052	2	US-08-790-912-2	Sequence 2, Appli
36	128	4.3	309	2	US-08-405-175A-8	Sequence 8, Appli
37	127.5	4.3	532	1	US-08-285-440-5	Sequence 5, Appli
38	127.5	4.3	532	1	US-08-630-349-5	Sequence 5, Appli
39	127.5	4.3	1360	4	US-09-788-657-22	Sequence 22, Appli
40	127	4.2	1002	4	US-09-252-991A-27980	Sequence 27980, A
41	127	4.2	1269	4	US-09-645-456A-15	Sequence 15, Appl
42	127	4.2	1269	4	US-09-435-324A-15	Sequence 15, Appl
43	127	4.2	1269	4	US-09-645-791-15	Sequence 15, Appl
44	127	4.2	1298	4	US-09-645-456A-14	Sequence 14, Appl
45	127	4.2	1298	4	US-09-425-324A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-051-019-2

Sequence 2, Application US/09051019

Patent No. 6103229

GENERAL INFORMATION:

APPLICANT: KAHMANN, Regine and QUADBECK-SERGER, Claudia

TITLE OF INVENTION: Regulatory gene from Ustilago maydis

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Keil & Weinkauff

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage

COMPUTER: IBM AT-compatible, Pentium processor

OPERATING SYSTEM: Windows 98

SOFTWARE: WordPerfect version 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/051,019

FILING DATE: 31-MAR-1998

CLASSIFICATION:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2289 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-09-051-019-2

Query Match 5.8%; Score 174.5; DB 3; Length 2289;

Best Local Similarity 20.1%; Pred. No. 5.2e-06;

Matches 120; Conservative 85; Mismatches 184; Indels 209; Gaps 28;

QY 28 GFRITTIAXREGAGDADADPLEPPPPQOGLSLRSGRQPKETVEQVPLTIARRGR 87

DB 1387 GSRLLTADNR---DRELD---KLVERVEDATDPADQKP 1420

QY 88 SMPVLSDEGPTSCPATDAETASGVSR---SASTRSGPOS---AS 129

DB 1421 NANCICRSMPIAIPSSSGAECRCRVYHLSCKIVRSEVSRAGGWVCPFCWYGSA 1480

QY 130 TAVKERPASS---EKVKGDDH---DDTSDSDSDGHTLKLQNLRRKREQETEPKLG 183

DB 1481 PFLKREKATSIADLSKLVYDQDHRDQFKFLPLEWDAISEVV--AKAKGFETAAKRMKT 1538

QY 184 IQRLRKRRREE-----GPAS-----TVGSEAST----- 208

DB 1539 L-SLARDDQKVILAWLRRSGICPVDTLGPKEVNMMLISENLLALGSGQGGAAAPV 1597

QY 209 --VSGVLPKOB-----PENDQGVVSOAGKDDRESKLE-GKAAQDIK 247
 DB 1598 ERKASTPARSDERTTTPLPRSSRPAPADDSG--SPAVDDRRKAKRKRKAKLP 1655
 QY 248 DERPGDLGRPKPECEGVDNALYICICROPHNNFMICCDCEWFGDCVGVSEARGLL 307
 DB 1656 QEEIG-IGAYR-----DRPIYCLCHEPESGR-MTACDKCMLMFWTNCVRLDDPP-- 1703
 QY 308 ERNGEDVICNCTILOVDETHSTADQO--EAKWRPGDADGTD-----CYSIGTIHQ 358
 DB 1704 NLGNRPICPMCCIT-----KAEKRYFOAEVRVKGIDGVPDPLMLDIRATLSLEK 1753
 QY 359 KSSRDQGIKRIEKAAPSGKK--KLKIFQPIVEAPGASKICGPGCHVAPQDSVYCSN 415
 DB 1754 PVS-----KIQSWTSPENKRVILVLEKETPAIHA----- 1782
 QY 416 DCILKHAATMKFLSSCKOKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 475
 DB 1783 -----EEVHSQITKQARLESPTPSKARVSLGRSDSIST---PA---K 1818
 QY 476 ETTVKAAPVVPARSEALG-----KEAACSSSPSWASDENYNAVXP---EKTAAPSP 524
 DB 1819 ESGAVPYAAPVPEAGVIGVLPALTPAADSPASRGNDSDFAAASPPLWDAKTGSP 1876
 RESULT 2
 US-09-252-991A-18387
 ; Sequence 18387, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 18387
 ; LENGTH: 1075
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-18387
 Query Match 5.5%; Score 164.5; DB 4; Length 1075;
 Best Local Similarity 21.5%; Pred. No. 1.4e-05;
 Matches 126; Conservative 76; Mismatches 228; Indels 155; Gaps 26;
 QY 7 PSNEEAPKAKPTSKPRKWTGFRRTTIAKREGAGDAEADPLE--PPPPQQQLGLSLERS 64
 DB 553 PQTEAPVPEAKPMPSPSLFQGLVKSIVGLFACKQPAKPARTSKPAERQTRQDERN 612
 QY 65 GRQPKRTERVEQFLTIARRGRSMFVSLDSEPTSCPATDAHTASGVSASHTSG 124
 DB 613 GQONRRDGRD---GNRRDERKPREERAERQPRB-----ERAERNREERSRRR 662
 QY 125 PQSASTAVKRPASSKRVKSGDDHDDTSDSDGLTKELQNLRLKREQRPTEH---P 180
 DB 663 ERA-----ERPAERB-----QPREGREERAEHTPREERQ 693
 QY 181 LKGIQSR-LKKREEREPARTVSEASDVTVEGLPSKQEPENDQGVVSOAGKDDRESK-- 237
 DB 694 REGREGRERSEHRRERARERARERARERARERARERARERARERARERARERAR 749
 QY 238 --LEGKA-----AQIDKBBPGDILGRPKPECEGYDPNLYCICROPHNNFMICCDRC 288
 DB 750 AALEABALPNDESLEQDDQDDTGE--RPERRSRQ-----RRRSNR--ERQ 793
 QY 289 EWFHGDVCISEARGKLLRENSDICYNCTI-----LQVDEB-----THS 330

DB 794 RE-VSGVEG-SEAT-----DAAAPLNTVAAAAAGIIVASEAVEANVQAPATTS 843
 QY 331 ETADQOQAKRMPGDADGTDCTISGTTIRKSSSEQOQIKGR-----IEKAANPSGKK 381
 DB 844 EAASETTA-----SDTASTSEAVTQDADSEANTGETADIEAPVTVSVWRDEADQST 897
 QY 382 LKIQPIVIEAPGASKCI-----GPGCHVAPQDSVYCSNDCILK 420
 DB 898 LLVQAQATEEAPFASGVSSESDAESAQVAPATEAAEVAAPVVEVAAPSFAATEEPTPA 957
 QY 421 HAA-----ATMKFSLSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 476
 DB 958 IAAVAPANATGALNDPREKRIQREAREAREAREAREAREAREAREAREAREAREAREARE 1014
 QY 477 TTVKAVVVPARSALKEACSSSPSWASDENYNAVXP---EKTAAPSP 521
 DB 1015 ASAQEE---PAAPQA---EBITQADVPSQA-DEAQAQVAPPEAS 1052
 RESULT 3
 US-08-769-309A-5
 ; Sequence 5, Application US/08769309A
 ; Patent No. 5741890
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klauck, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/769,309A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5741890and, Greta E.
 ; REGISTRATION NUMBER: 35,302
 ; REFERENCE/DOCKET NUMBER: 27866/33451
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-474-6300
 ; TELEFAX: 312-474-0488
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1780 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-769-309A-5
 Query Match 5.4%; Score 161.5; DB 1; Length 1780;
 Best Local Similarity 19.4%; Pred. No. 5.1e-05;
 Matches 133; Conservative 76; Mismatches 221; Indels 255; Gaps 28;
 QY 3 DKGDPSNEEAP--KAKNPTS-----KEFRKTW-GFRRTT----- 33
 DB 266 EGEKE 325
 QY 34 -----IAKREGAGDAEADPLEPP----- 384
 DB 326 KKEQEPKVTERTDGRKAEVASEKLITASEQAHQDPAESAHEPRLSAEYKVELPSEQV- 384

309 RRGEDYICPMCTILQVQDETHSETAQOQAKWPGDADGTDCTSIGTIEKQSKSEDOGIGK 368
489 EGHPD---PDAELQMEQE---AERRQPOIKQEPSEEE---ERQKREKRE 535
369 RIEKAAMPKGLK-IPQVIEAPGASKICIGPCCHVAQPDVSYVCSNDC---ILKHAAT 425
536 PHEEEEPKQCLKPTLRPISSAPSVSSASGN---ATENTPGDESPOGIIIPHE--- 587
426 MKFLUSSKQKQP---KPKKMKK---PEKPSLPKCAQAGIKLSSVHKPAPKKTIV 479
588 ---NSPDQOQPEHPRKPKIGLSLKLGSASNPGQFNSVTKRKLFPVDSVFNKPFEDSDSDVP 643
480 KKAUVVP 486
644 RRRKLVP 650

RESULT 7
US-08-801-308-1
; Sequence 1, Application US/08801308
; Patent No. 6368790
; GENERAL INFORMATION:
; APPLICANT: Scott, Robert E.
; TITLE OF INVENTION: cDNA ENCODING P2P PROTEINS AND USE OF
; TITLE OF INVENTION: P2P cDNA-DERIVED ANTIBODIES AND ANTISENSE REAGENTS IN
; TITLE OF INVENTION: DETERMINING THE PROLIFERATIVE POTENTIAL OF NORMAL,
; TITLE OF INVENTION: ABNORMAL AND CANCER CELLS IN ANIMALS AND HUMANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Weiser & Associates, P.C.
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,308
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.6435P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-801-308-1

Query Match 5.3%; Score 157; DB 4; Length 1404;
Best Local Similarity 20.5%; Pred. No. 9.1e-05;
Matches 133; Conservative 77; Mismatches 228; Indels 212; Gaps 30;
8 SNERAPK--AKPTSKFKTKMGFPRTTIAKRBGAGDADPLEPPPPQQLGLSLRRSG 65
851 AKKATKIDSVKPSSSQKD---EKVGTGPKAKSKAKDTRRSQPR-----TRRSK 900
66 R-----QPKKTERVQPLTI-----ARRGRSMFVLSDSG-----RPT 100
901 RTVPKTSQKQVATRTPRSLKINYLAREKREKREKSKVDKDFSSSMKSKVGT 960
101 SCPATDAETASGVSASSETSGPQASATAVKERPASSEKVKGGDDHDDTSDSDSGIT 160

CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-257-179-80

Query Match 5.4%; Score 161; DB 4; Length 238;
Best Local Similarity 33.7%; Pred. No. 3.3e-06;
Matches 33; Conservative 17; Mismatches 18; Indels 30; Gaps 4;
230 GKDRESKL-----EGKAQDIKDEPDGLGRPKPECEGYDPNALYICRQPHN-NRF 281
32 GITEKAKMDVYICNDCEAGQGSSE-----LYCICRTPYDESQF 73
282 MICDRCCEWFGDCVGISEARGRLRNGEDYICPNC 319
74 YIGCDRCQWYGRGCVGLQSAELT-----DEYVCPQC 107

RESULT 6
US-10-164-595-56
; Sequence 56, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 56
; LENGTH: 779
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-56

Query Match 5.3%; Score 158.5; DB 4; Length 779;
Best Local Similarity 21.4%; Pred. No. 2.9e-05;
Matches 104; Conservative 58; Mismatches 191; Indels 134; Gaps 19;
20 SKKFRKTFRTTIAKRBGAGDADPLEPPPPQQLGLSLRRSGRQPKRTERVQPLT 79
278 SREISK---FRDTHKKLEKGGKKE---RQIEKRRERERE 315
80 IARRRRRSMFVLSDSGSEPTSCPATDAETASGVSASSETSGPQASATAVKERPASS 139
316 RERERER-----ERERERERERERERERERERERERERERERERERER 357
140 EKVGGDDHDDTSDSDGLTLKELONLRKRQEQETPLKIGIQRLLK----- 191
358 EREERDKKRDDEDEDAYERKLEKRLBK-BAAYQER-LQWIRKRTKTRVETKKA 415
192 RREEGPAETVGEA---SDTVGVLPKQBPENDQGVVQAGDDRESKLEKGAADIKD 248
416 EREERERREKAKKALKEGLEDYDDDDPKYRGSAQKRLRDREKMEA-----D 468
249 EEPGLGRPKPECEGYDPNALYICRQPHNRFMICDRCCEWFGDCVGISEARGELLE 308
469 ER--DRKREKELE-----EIRQLLIA 488

Db 961 EIVKPSKRWGVDGVEKLETT---PEKDKIASSTTPA---KKIKLNR-----T 1003
Qy 161 LKELONLRKRKEQPTFRPLKGIQSL-----RKRRREGPAETV----- 201
Db 1004 GKIKGNAENASTTKPSEK-LESTSSKIKORVKGKAKRVAGSESSSTLVDTSTST 1062
Qy 202 -----GSEASDVTVEGLPSKORPEND-----QGVVSOAGKDDRESKLEGA 242
Db 1063 GGSFVRKSEKTDKRTVINTMBEYNNNTAPAEADVIIMIIVPQSKWDKDFSEEB--- 1119
Qy 243 AQDIKDEBP-GDLGRP-----KP-----BCEGYDPNLYCICRQPHNRFWICC 285
Db 1120 --DVKTTPQTSVGKPSIIFKNVTKFSATAKYTEKSEOPKLOKLPKEASH----- 1170
Qy 286 DRCEWPHGDCVGSSEARGLLER--NGEDYICPN-----CTILQVODETHSET 332
Db 1171 ELWQHELRSKGSNSSEKRAKREHSGSEKNDPKKSGNQPKBSTVDLSQGHFKT 1230
Qy 333 ADQOEAKWRPGDAGTDTCTSIGTIBQKSDGIGIKRIEKAANPSGKKUKIFQPVIEAP 392
Db 1231 LSGSSKETR-----TSEKHESVRGSSNKDFTFGRDKVD----- 1264
Qy 393 GASKCIGPGCHVAQPSVVCNSDCILKHAATMKFLSSGKQKPKKB-----KMKMKP 447
Db 1265 -----YSDRYSSKRDDEGELARRDSPPRKESLSQKSKIRE 1305
Qy 448 EKPSLPKCGAAGIKTSV--HKRPAPEKKTETTKGAVVVPARSEALGKBAACESPTPSW 505
Db 1306 ER-DLPKGAESKKSNSPPRDKKPHKAPYETKR-----PCEETKP-- 1347
Qy 506 ASDHYNNAVPEKTAAPS-----BSLLYKVMVHLGVGLDPSRSPWIAIPW 551
Db 1348 -VDKN-SGKERKHAERNGKSSGANCHVILYTRQL-FWRSSWLLGRW 1394

RESULT 8

US-10-164-595-80
; Sequence 80, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 80
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-80

Query Match 5.1%; Score 152.5; DB 4; Length 735;
Best Local Similarity 20.8%; Pred. No. 9e-05;
Matches 108; Conservative 75; Mismatches 227; Indels 109; Gaps 20;
Qy 2 DDKGDSNEAPKAIKPTSKFRTWGR---RTTIK---REGAGDAEADPLEPPPPQ 54
Db 163 EDKRLISREISK-FRTHKKLEEGKKEKEQIEKERRERERERERERERERERER 221
Qy 55 QQLGLSLRRSGRQPKRTERVEQFLTIARRGR-----RSMFVSLDSGSEPTSCPATDA 107
Db 222 EREREREKERERERERDRDRTKDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR 281
Qy 108 ETASEGSVSASSTRSGPOSASTAVKRPASSEKVGCGDDHDDTSDSDGLTKELQNR 167
Db 282 DRER 341
Qy 168 LRKREQEPTEPRPLKGIQSLRK---RRREGPAETVGEA---SDTVEGVLPSPK 216
Db 342 LREK-EAAYQER-LKNWIERKKTVEYEKEAREERERERERERERERERERERER 399

Qy 217 QRPENDQGVVSOAGKDDRESKLEGAQAQDIKOBEPGDLGRPKPCEGYDPNLYCICRQP 276
Db 400 DDPKYRGSALQKRLADREKMEA-----DER--DRKREKEELE----- 436
Qy 277 HNRPMICCDRCBETHGDCVGSSEARGLLERNGEDYICPNCTILQVODETHSETADQ 336
Db 437 -----BIRQLLAEGHPD-----PDAELQRMBOB--AERRRQP 467
Qy 337 EAKWRPGDADGTDCTSIGTIEOKSSSDGILKRIEKAANPSGKKLK-IPQPVIEAPGAS 395
Db 468 QIKQEPSESEEB-----EKQEKERERERERERERERERERERERERERERER 519
Qy 396 KCIQPGCCHVAQPSVVCNSDC--ILKHAATMKFLSSGKQKRP---KPKKKKKMK--P 447
Db 520 SASGN-----ATENTPDESFCGIIIPHE-----NSPDQOQPEHRPKIGLSLKIGAS 567
Qy 448 EKPSLPKCGAAGIKTSVVCNSDCILKHAATMKFLSSGKQKRP---KPKKKKKMK--P 447
Db 568 NSPQGPNSVKRKLVPDSVFNKFEDESDDDVPRKRKLVP 606

RESULT 9

US-10-164-595-79
; Sequence 79, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 79
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-79

Query Match 5.1%; Score 152.5; DB 4; Length 784;
Best Local Similarity 20.8%; Pred. No. 9.9e-05;
Matches 108; Conservative 75; Mismatches 227; Indels 109; Gaps 20;
Qy 2 DDKGDSNEAPKAIKPTSKFRTWGR---RTTIK---REGAGDAEADPLEPPPPQ 54
Db 212 EDKRLISREISK-FRTHKKLEEGKKEKEQIEKERRERERERERERERERERER 270
Qy 55 QQLGLSLRRSGRQPKRTERVEQFLTIARRGR-----RSMFVSLDSGSEPTSCPATDA 107
Db 271 EREREREKERERERERDRDRTKDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR 330
Qy 108 ETASEGSVSASSTRSGPOSASTAVKRPASSEKVGCGDDHDDTSDSDGLTKELQNR 167
Db 331 DRER 390
Qy 168 LRKREQEPTEPRPLKGIQSLRK---RRREGPAETVGEA---SDTVEGVLPSPK 216
Db 391 LREK-EAAYQER-LKNWIERKKTVEYEKEAREERERERERERERERERERERER 448
Qy 217 QRPENDQGVVSOAGKDDRESKLEGAQAQDIKOBEPGDLGRPKPCEGYDPNLYCICRQP 276
Db 449 DDPKYRGSALQKRLADREKMEA-----DER--DRKREKEELE----- 485
Qy 277 HNRPMICCDRCBETHGDCVGSSEARGLLERNGEDYICPNCTILQVODETHSETADQ 336
Db 486 -----BIRQLLAEGHPD-----PDAELQRMBOB--AERRRQP 516
Qy 337 EAKWRPGDADGTDCTSIGTIEOKSSSDGILKRIEKAANPSGKKLK-IPQPVIEAPGAS 395
Db 517 QIKQEPSESEEB-----EKQEKERERERERERERERERERERERERERERER 568
Qy 396 KCIQPGCCHVAQPSVVCNSDC--ILKHAATMKFLSSGKQKRP---KPKKKKKMK--P 447

Db 569 SASGN-----ATPNTPGSGCGIIPHE-----NSPDQOQPEHRPKTGLSLKLGAS 616
Qy 448 EKPSLPKCGAAGIKISSVHRKPAPEKKEKTTVKKAVVVP 486
Db 617 NSPGQPSVXKELPVDVSFNKPEDESDDDVPRKRLVP 555
RESULT 10
US-10-164-595-54
; Sequence 54, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 54
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-54
Query Match 5.1%; Score 152.5; DB 4; Length 843;
Best Local Similarity 20.8%; Pred. No. 0.00011;
Matches 108; Conservative 75; Mismatches 22; Indels 109; Gaps 20;
Qy 2 DKGDPSENEAPKAIKTSKFRKTWFR-----RTTIAK-----REGAGDAEADPLEPPPPQ 54
Db 271 EDKDLISREISK-PRDTHKLEBKKGKKEKQBIKERERERERERERERERERERERERER 329
Qy 55 QQLGLSLRSRSGQPKTERVQFLTIARRGR-----RSMVSLSDSGEPTSCPATDA 107
Db 330 EREREREKERERERERDRDRDKERDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR 389
Qy 108 ETASGSESVASSETSGSQSASTAVKRPASSEKVGKGGDDHDDTSDSDGLTKELQNR 167
Db 390 DRER 449
Qy 168 LRRKREQEPTERPLKIGOSRLKK-----RREGPAETVGSBA-----SDTVEGLPSK 216
Db 450 LREK-EAAYQER-LQNWIREKTKTREYKEAREEREREREREMAXKGLKFLDYDDDR 507
Qy 217 QEPENDQGVSAQKDDRESKLEKGAQDIKDEEPGLGRPKPECEGYDPNALYICRQP 276
Db 508 DDPKYRGSALQKRLDRKEKEA-----DER--DKREKEHE-----544
Qy 277 HNNRPWICCDRCBWFHGDVCGISARGRLLRNGEDYICNCTILOVDTHSETADQ 336
Db 545 -----BIRQLLAEGHPD-----PDASELQRMQE-----AERRRQP 575
Qy 337 BAKWEPGADGDTCTSIGTIRKSSSEDOIGRIEKAANPSKKLK-IPQVTEAPGAS 395
Db 576 QIKQPESEHEE-----BEKQEKREPEEMEEPEEPQKPKLKTLPISAPSVS 627
Qy 396 KCIQGCCHVAPQDSVYCSNDC--ILKHAAATFMKFLSSGKEQKP---KPKERKMK---P 447
Db 628 SASGN-----ATPNTPGSGCGIIPHE-----NSPDQOQPEHRPKTGLSLKLGAS 675
Qy 448 EKPSLPKCGAAGIKISSVHRKPAPEKKEKTTVKKAVVVP 486
Db 676 NSPGQPSVXKELPVDVSFNKPEDESDDDVPRKRLVP 714

RESULT 11

US-08-978-277A-4

; Sequence 4, Application US/08978277A

; Patent No. 6582956

; GENERAL INFORMATION:

; APPLICANT: Geiman, Irwin H.

; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/978,277A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: 08/665,401
; FILING DATE: 18-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S
; REGISTRATION NUMBER: 26,154
; REFERENCE/DOCKET NUMBER: A30558 - 165/34008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2558
; TELEFAX: 212-765-2519
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1596 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-978-277A-4

Query Match 5.1%; Score 152.5; DB 4; Length 1596;

Best Local Similarity 19.6%; Pred. No. 0.00027;

Matches 141; Conservative 78; Mismatches 236; Indels 263; Gaps 32;

Qy 2 DKGDPSENEAPKAIKTSKFRKTW-GFRRTTIAK-----EGAGDA 43
Db 270 ESPSPVNSSETTSFK---KFFTHGAGWKTSFKSKEDDLETAERKEQAEKVDDE 326
Qy 44 RADPLEPP-----PPQOLGLSLRSRSGQ---PKETERVEQ 76
Db 327 EKEKTEPASEEQPFAEDTQARLSADYKVELPLDQVG-DLEASSEKCAPLATFVDE 385
Qy 77 FLTIARRGRRRSPVSLSDSGEPTSCPATDAE--TASEGSVES-----ASSTRSGPOSAS 129
Db 386 KME-AHQEVVAHVHVSIVKTEBEEQGGGAEQGVVVEGTGSLPPKLAEPQVPOEAE 444
Qy 130 TAVKERPASSEKVKGGDDHDDTSDSDGLTKLE-----LQNRLEKKEQEPTE 178
Db 445 PA-BELMKSRMCMVSGGDHTQLTDLSPERKTLPKHPGIVSEVEMLSQERIKVQGSPLK 503
Qy 179 R-----PLKGIQSRLEKRR---EEGPAB-----TVGSASD-----207
Db 504 KLFSSGGLKLSGKKQKGGGGDEEPGEVQHHTSPESADQKGGSSASSPEEPET 563
Qy 208 --TVEGVLPKQEPENDQGVVVSQAK-----DRESKLE 239
Db 564 TCEKGPLEAPQDGEABEGITSDGEKREKRGITPWASFKKMVTPKKVRPRPSESDEKELE 623
Qy 240 -----GKAAQDIKDEBPGLRPKPECEGYDPNALYICIKQPHNRFMICCRCE 289
Db 624 KVSATLSSTSTVSEMQDEVKTVGEEKQE-----EPKRR-----VDTSV 664

```

QY 333 ADQBEAKVRGCDAG--TDCTSIGTIEOKSSD-----OGIKGRLEKKAANPSGKK-- 381
Db 973 EDESASK---AEADAVIRKRVSVASDDRAEDMDAIEKGAEQAQESEADEBKEDA 1029
QY 382 -----LKIPQPVIEAPGASKICJ-----PGCHVAQPDSVY 412
Db 1030 REEYZEPKEHAEADYYMAVVDKAAEAAGABEQYGLFTPTTKQLCAQSPG----- 1078
QY 413 CSNDCILKHAATW--KFLLSGREQK--PKPEKMKMKPEKPSL.PKCGAQAGIKISSVHK 468
Db 1079 -----REPASSIHDTLPGGSESEATASDENRRDPPEFTATSYGTOSTHEISS--- 1128
QY 469 RPAPEKKEKTTVKAAVVVPARSEALGKEACESTPSS 504
Db 1129 EPTFMDMSTPRDWMSDETNNE-----ETESPSS 1156

RESULT 13
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; US-08-714-741-32

Query Match          5.0%; Score 149.5; DB 4; Length 8991;
Best Local Similarity 20.5%; Pred. No. 0.0056;
Matches 129; Conservative 73; Mismatches 260; Indels 167; Gaps 25;

QY      8 SNEAPEALKAPTS-----KEFRKTGWFRRTTAKREGAGD-----AEADLEPPPPQQOL 57
       : :: ||| |   :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-464LAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-953-913-2

Query Match      5.0%; Score 148.5; DB 1; Length 1805;
Best Local Similarity 19.8%; Pred. No. 0.00072;
Matches 132; Conservative 105; Mismatches 223; Indels 207; Gaps 36;

QY      10  EAPKAKTPTSKFRKTWG--FRRTIAKRGAG-----DRAADPLEPPPPQOQLG 58
Db      521  EETQDQGFLQKTKALGEELMSLKIQNYETAGKNCNSTEGHGLTLEGPEKEKQIP 580
QY      59  L-SLRSS-----GROPKPTE-----RVEQFLTIAR-----82
Db      581  LKSLSEKNVSEKTLNGVPISELLGKGDTRTDOELMSPKGTILKRFSSLGKSEQVVR 640
QY      83  --RGR-RSMPEVLEDSGPTSCPATDAATASEGSVBSASETRSGPOSASTAVKE--RPA 137
Db      641  PSKEGNLESTAKESQHPGLPPGAE-QDMLERLVE--KEQDSFPRSPBEDQACRPL 697
QY      138  SSEKVGGDHDDTSDSDGLTKLQNLRRKREQBETPEAPLKGIOQLRKYGRBEP 197
Db      698  QKE-----NQEPILGYEAEQ---QILRLIKESQESILRSPFEEDQAGSLQKEN--745

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:51:17 ; Search time 45 Seconds
(without alignments)
3940.472 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGPSENEAPKAIKPTS.....RSFWIAIPWACGLGVAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329.5	11.0	2016	5 Q9VG78	Q9VG78 drosophila
2	217	7.3	663	5 Q9W352	Q9W352 drosophila
3	202.5	6.8	366	5 Q8T3Y1	Q8T3Y1 drosophila
4	197	6.6	2649	5 Q9W0T2	Q9W0T2 drosophila
5	197	6.6	2669	5 Q9SVB8	Q9SVB8 drosophila
6	197	6.6	2669	5 Q9W0T1	Q9W0T1 drosophila
7	194	6.5	3146	5 Q9VUB5	Q9VUB5 drosophila
8	191	6.4	2764	4 Q7Z7D6	Q7Z7D6 homo sapien
9	187.5	6.3	542	5 Q91870	Q91870 euplates cr
10	187	6.3	443	11 Q8C969	Q8C969 mus musculus
11	187	6.3	473	11 Q8C9E0	Q8C9E0 mus musculus
12	187	6.3	17352	5 Q9SYM2	Q9SYM2 procamburus
13	182	6.1	563	13 Q7SZK6	Q7SZK6 brachydanio
14	182	6.1	563	13 Q7SZK4	Q7SZK4 brachydanio
15	181	6.1	513	5 Q95P08	Q95P08 chironomus
16	181	6.1	720	5 Q7Z2C9	Q7Z2C9 drosophila

17	178	6.0	606	6	Q28687	Q28687 oryctolagus
18	178	6.0	2400	4	Q8IWP2	Q8IWP2 homo sapien
19	178	6.0	2416	4	Q8IWP1	Q8IWP1 homo sapien
20	178	6.0	2432	4	Q8IWP0	Q8IWP0 homo sapien
21	178	6.0	2448	4	Q8IWN9	Q8IWN9 homo sapien
22	178	6.0	2464	4	Q8IWN8	Q8IWN8 homo sapien
23	178	6.0	2480	4	Q8IWN7	Q8IWN7 homo sapien
24	177	5.9	2464	4	Q86SQ1	Q86sq1 homo sapien
25	176.5	5.9	1417	3	Q871Y7	Q871y7 neurospora
26	175	5.9	424	3	Q74508	Q74508 schizosacch
27	175	5.9	545	11	Q8VDN7	Q8vdn7 mus musculus
28	175	5.9	5327	5	Q76891	Q76891 drosophila
29	174.5	5.8	669	11	Q35540	Q35540 mus musculus
30	174.5	5.8	687	11	Q9QX75	Q9qx75 rattus norv
31	174.5	5.8	2289	3	Q9HFW4	Q9hfw4 ustilago ma
32	174	5.8	5412	5	Q9W596	Q9w596 drosophila
33	172	5.8	1135	6	Q8HZN3	Q8hzn3 canis famil
34	171.5	5.7	700	4	Q9BUV3	Q9buv3 homo sapien
35	171.5	5.7	1110	13	Q91255	Q91255 petryozon
36	170	5.7	510	5	Q45407	Q45407 caenorhabdi
37	170	5.7	878	13	Q8AVW4	Q8avw4 xenopus lae
38	169.5	5.7	620	10	Q8LNZ4	Q8lnz4 nicotiana t
39	169	5.7	578	11	Q9JL92	Q9j192 mus musculus
40	169	5.7	590	13	Q91803	Q91803 xenopus lae
41	169	5.7	1893	11	Q8CJ14	Q8cj14 rattus norv
42	169	5.7	4969	11	Q8CF91	Q8cf91 mus musculus
43	169	5.7	5165	11	Q8CF92	Q8cf92 mus musculus
44	167	5.6	1408	5	Q9W0C9	Q9w0c9 drosophila
45	166.5	5.6	461	11	Q9QX76	Q9qx76 rattus norv

ALIGNMENTS

RESULT 1

Q9VG78 PRELIMINARY; PRT; 2016 AA.

AC Q9VG78; (1-MAY-2000 (TREMBLrel. 13, Created))

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE CG6525 protein.

DE SPP OR CG6525.

OS Drosophila melanogaster (fruit fly).

OC Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

[1]

SEQUENCE FROM N.A.

RC STRAIN=Berkely;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,

RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Beres P.V., Beruan B.P., Shandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,

RA de Pablos B., Belcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Flesschmann W.,

RA Roelber C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein D.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
DR FMBL; AE003695; AAF54807.1; .
DR FlyBase; Fg0003604.1; Spp.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006576; BRK.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00592; BRK; 1.
DR SMART; SM00245; PHD; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS00166; ZF_PHD 2; 1.
SQ SEQUENCE 2016 AA; 221998 MW; FB61A332EA710F5E CRC64;
Query Match 11.0%; Score 329.5; DB 5; Length 2016;
Best Local Similarity 22.1%; Pred. No. 5.1e-14;
Matches 159; Conservative 73; Mismatches 220; Indels 269; Gaps 29;
QY 8 SNEAPKAIKTSKFRKTMGTGRTTIKRGAGDAEADPLE-----PPPPQQLG---- 58
DB 639 ADEDIPKLVKLEST-----AADLPPVPVSLPTNIPAPMQGVGNA 681
QY 59 -----LSLRSGRQPKRTVEQFLTIARRGRSRSPVSL 94
DB 682 AADPNIMDTANEIDTADFLQHVGLIEDQK-PEAEVVKVLA-STEPGTLDAIVMP 739
QY 95 DSGEPTSCPATDAETASGV-ESASETRSGQASST-----AVKERPASSEKVKGGDDH 148
DB 740 TSIEPVDVQAHNLLNENASLETPAQSMSTLSIACSTPSRVAASPTPTSAKVRG---- 795
QY 149 DTSDSDSLGLTKELQRLRKREQEPTEPLKGIQSLRKKREE-----GPATWGS 203
DB 796 -----YGRVILYLPPIEAP-----TTRAKRRAQPPSPAPMAATSS 830
QY 204 EA-----SDTVGVLPSKQEPENDQGVVSQAGKDDRESKLE 239
DB 831 DAGNLSPGSSSLDASINQPLNTSSLNDSPQSGPKRPNP-REPSMARRESTAPRSKKLD 889
QY 240 GKAAOD---IKDEEPDGLRPRCEGYDPNLYCICROPHNNRPMICCDREWFHGD 296
DB 890 ASQNDPDASSEQEDD-----DPNKLWCICROPHNNRPMICDCEDFHGTG 938
QY 297 VGI SARGLLERNRGEYICENTILQVQDETHSETADQQRKWRPGDAGTDCSTIGTI 356
DB 939 VGVTRKAGTDMENKGLDWKCPKC-----VKQERSQRIIDMLVTRPTTP 985
QY 357 EQKSGEDQGI--KGRIEKAANTSG-KKKLKIFQPVTRAP-----GASK---- 396
DB 986 EQRPSETKVLTTTARIQVQAAPSAPRTLPVVLTVASSPMRIAPAKPKKFTGAISHQ 1045
QY 397 -----CIGPG-----CCHVAQPSVYCSNDCILKHA----- 422
DB 1046 QOQLNFIKGPFGKRIISTLCVWCRPASTSSVYGCICIRKYSQAIOAHATKGPLP 1105
QY 423 --AATMKFLSSGKQKPKPKKX----- 443
DB 1106 QNAGAQLNLNNGFPAKKNKKDLFDVLRLQADTVSKVERINVERKSGRVITGEMAPSAH 1165

QY 444 ---KMKPEKPS---LPRCGAAGIKISSVHKR---PAPKKETTVKAVVVPARS--BAL 492
DB 1166 QFRKWLQENPSFEVLPSGTQVS-----ADAERKLKGAPEAATSTSEPAVLGVAKPPEGP 1221
QY 493 GKKAACSSSPSWASDNNYNAVRP-----EKTA-----APSPSLLYKCMVHLGVGLDPS 542
DB 1222 AKUSHPONTTVQASHQIGISVRLPKXDKKEKTIPTVQATPN-----RIAGKPEPV 1274
QY 543 R 543
DB 1275 R 1275
RESULT 2
Q9W352
ID Q9W352 PRELIMINARY; PRT; 663 AA.
AC Q9W352;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CGI7446 protein.
GN CGI7446.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Ryans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Pfeiffer B.D.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Brattier P.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Chetty K.C., Bueam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idagwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein D.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
DR FMBL; AE003446; AAF46483.1; .
DR FlyBase; Fg00030121; CGI7446.
GO; GO:0003677; F:DNA binding; IEA.

[illegible]

STRAIN-Berkeley;
 MEDLINE-20196006; PubMed-10731132;
 Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benton P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 Burris K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 Foster C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris D., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski W.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 [3]
 SEQUENCE FROM N.A.
 Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 Bamron J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,
 Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,
 McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of *Drosophila melanogaster* genome";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 Mixra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 Kronmiller B., Marshall B., Millburn G., Richer J., Russo S.,
 Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 "Annotation of *Drosophila melanogaster* genome";
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [6]

DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS0014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2649 AA; 298507 MW; COD7F7E015EA5403 CRC64;

Query Match 6.6%; Score 197; DB 5; Length 2649;
Best Local Similarity 19.0%; Pred. No. 0.0001;
Matches 82; Conservative 75; Mismatches 139; Indels 136; Gaps 15;

Qy 49 EPPPPQOOLGLSRRGRQPKTER-----VEQFLTIARRGRRRMPVSLD----- 95
Db 2197 EHDEPTNLGLDISETDLENKQNESFVVTGRYIQKSISNALKQ--NLSPELEKLVCMQ 2254
Qy 96 SGEPTSCFATDAETASGSVESASSETSGQSASTAVKRPASSEKVKGGDDHDDTSDSD 155
Db 2255 KQENANSTNEWETCSRGVN-----EALTPSRQTDDETKIR-----TSLR 2298
Qy 156 SDGLTLKELQNLRRKREOEPTERPLKGIQSRKLRKRRREGPAETV-----GSE 204
Db 2299 PNAMTSSQFNRLKKNRSKNDEVAELGEQKQSLERHKKLLKNILKRSLERLNLQSE 2358
Qy 205 ASDTVEGVLPKQEPENDQGVVQAGKDDR-ESKLEKGAQDIDKEPGD----LGRPKP 259
Db 2359 IHEDVTKVQVRHVRP-----LSNASPDEQSENERSGEPNLDPRTEVQNPFRHGAGRPKK 2412
Qy 260 -----ECE-GYD 265
Db 2413 LTRKKEKLYCICRTFYDDTKFYVGCDCLSNWFHGDVCSITEASKLSEFICIDCKRARE 2472
Qy 266 FNALYCICROPHN-NRFNICDCRCEWFGHDCVGISEARGELLERNGEDYICNCTILOV 324
Db 2473 TQQLYCSQROPYDSQFYICDKQDFHGRGVGLQSEAEFI-----DEVVCPECO----- 2524
Qy 325 QDETHSETAQOERAKWRPGDAGTDCSTIGTIEQKSSDDQIKGRIEKAA----NPSGKK 380
Db 2525 -----KNDANANMKKLTNDVVELKNLIKQMLKQHLKSAWPFPEVDPK 2568
Qy 381 KLKIFQPVIEAP 392
Db 2569 EAPDYKVIKEP 2580

RESULT 5
Q95VB8 PRELIMINARY; PRT; 2669 AA.
AC Q95VB8;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Nucleosome remodeling factor, large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21468388; PubMed=11583616;
RA Xiao H., Sandaltzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
FA Fu D., Wu C.;
RT "Dual functions of largeest nurf subunit nurf301 in nucleosome sliding and transcription factor interactions.";

MoL. Cell 8:531-543(2001).
EMBL; AF417921; AAL16644.1; --
FlyBase; FBgn000541; E(bx).
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR00637; AT hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 3.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300687 MW; 6B4925AFFF49D6F CRC64;

Query Match 6.6%; Score 197; DB 5; Length 2669;
Best Local Similarity 19.0%; Pred. No. 0.0001;
Matches 82; Conservative 75; Mismatches 139; Indels 136; Gaps 15;

Qy 49 EPPPPQOOLGLSRRGRQPKTER-----VEQFLTIARRGRRRMPVSLD----- 95
Db 2217 EHDEPTNLGLDISETDLENKQNESFVVTGRYIQKSISNALKQ--NLSPELEKLVCMQ 2274
Qy 96 SGEPTSCFATDAETASGSVESASSETSGQSASTAVKRPASSEKVKGGDDHDDTSDSD 155
Db 2275 KQENANSTNEWETCSRGVN-----EALTPSRQTDDETKIR-----TSLR 2318
Qy 156 SDGLTLKELQNLRRKREOEPTERPLKGIQSRKLRKRRREGPAETV-----GSE 204
Db 2319 PNAMTSSQFNRLKKNRSKNDEVAELGEQKQSLERHKKLLKNILKRSLERLNLQSE 2378
Qy 205 ASDTVEGVLPKQEPENDQGVVQAGKDDR-ESKLEKGAQDIDKEPGD----LGRPKP 259
Db 2379 IHEDVTKVQVRHVRP-----LSNASPDEQSENERSGEPNLDPRTEVQNPFRHGAGRPKK 2432
Qy 260 -----ECE-GYD 265
Db 2433 LTRKKEKLYCICRTFYDDTKFYVGCDCLSNWFHGDVCSITEASKLSEFICIDCKRARE 2492
Qy 266 FNALYCICROPHN-NRFNICDCRCEWFGHDCVGISEARGELLERNGEDYICNCTILOV 324
Db 2493 TQQLYCSQROPYDSQFYICDKQDFHGRGVGLQSEAEFI-----DEVVCPECO----- 2544
Qy 325 QDETHSETAQOERAKWRPGDAGTDCSTIGTIEQKSSDDQIKGRIEKAA----NPSGKK 380
Db 2545 -----KNDANANMKKLTNDVVELKNLIKQMLKQHLKSAWPFPEVDPK 2588
Qy 381 KLKIFQPVIEAP 392
Db 2589 EAPDYKVIKEP 2600

RESULT 6
Q9W0T1 PRELIMINARY; PRT; 2669 AA.
ID Q9W0T1
AC Q9W0T1;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)


```

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.P.,
RA George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazell R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.M., Basu A., Bavendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ithegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Paclet J., Faragas V., Park S., Patel S., Pfeiffer B., Scheeler P.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler P.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Carlson B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL; AB003536; AAF49773.3; -
DR FlyBase; Fggn0036398; CG9007.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
SQ SEQUENCE 3146 AA; 336443 MW; 81EBAC5C308FEC31 CRC64;
Query Match 6.5%; Score 194; DB 5; Length 3146;
Best Local Similarity 20.5%; Pred. No. 0.00021;
Matches 121; Conservative 61; Mismatches 235; Indels 174; Gaps 24;
QY 33 TIAKRE-CAGDAEADPLEPPPPPOOQLGLSLRSRGRQKPTREVEQFLTIARRRGRSMP 90
DB 575 TIAKRSVSAPSRAVSLERKQHQQQLQHDVIGGR-KAPTIVIEY-----NKHGVNSIV 628
QY 91 VLSLGSSEPTSPATDAETASGSS-----VSSASETRSGPOSASTAVKRPASSSEKVG 144
DB 629 GSSNLLAQNSMSNLGAPRSGSGCFATTPATPLHLTPVNVFVHVEAAPSPSPALVK 688
QY 145 GDHDDTSDSDGLTLKQLNLRKREQEPT-RPLKGIQSLR-----189
DB 689 GSSPPPAQPOQQQQQHPGLPQMLNANDELIVIEVFPVPLTQDLRLOOLHAIMQDHTY 748
QY 190 -KKRKEGPAETVGSASDTVEGLVPSKQEP-----NDQVVSQAG-----230
DB 749 ASQOQQQQQPOQAAG---DTNPGAAQQVQPOQWSLGGIGVTVSGSQGTPTAVGYSYF 805
QY 231 -----KDRSEKLE-----GKAAQIDKEBPDLGRPKPCGSDYFNALYCIICROP 276
DB 806 GQCIASQADDDAHSASISSRMLASTDI-----DPGETETAPAEARSDSVTRCICELT 862
QY 277 HNNRFMICDRCEENFHGCVGISEARGRLIENGEDYICPNCITILQVODETHSETADQQ 336
DB 863 HDGTYMCCDKCSANQHVDGMLDR-----QWIPETWCELCPRAV-DKARALQRQ 915
QY 337 EAKWR-----PGDADTDTCTSTGTI-----EKSSEDQGIKGR1-----370
DB 916 KRKEHMLVATQAANGAAVAAGTTLSGGISGLPMSBELQHLASGLNGGPFATGMSK 975
QY 371 -----EKAANPSGKKKLIPOVIRAPGASKIGFCGCHVAQPSVYCSNDILKHAHA 424
DB 976 KSKTYKENGSTSLKTK-KSAVWGGEKNSGSG-----TPTG-----1014
QY 425 TWKPLSSGKEQKPKPKKKMKPKPKSLPKCAQAGIKLSSVHKRPAPKPKETTVKAVV 484
DB 1015 -----SSGKTSKSKSKSKSGD-----GSSGG-----GSSPALTAARKHAANLRQ---1056
QY 485 VPARSEALKEACSSSTPSWASDHNY---NAVKPKTA-----APSPILL 527
DB 1057 -----WIENYEATNHYSPFLRLHAIQKPSLL 1087
RESULT 8
Q727D6 PRELIMINARY; PRT; 2764 AA.
AC Q727D6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
```

01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Bromodomain PHD finger transcription factor.
 OS Homo sapiens (Human)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barak O., Lazzaro M., Lane W., Speicher D., Picketts D.,
 RA Shiekhattar R.;
 RT "Isolation of human NURF";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY282495; AAF22284.1; --
 SQ SEQUENCE 2764 AA; 308079 MW; 9785D7097C2099F9 CRC64;

Query Match 6.4%; Score 191; DB 4; Length 2764;
 Best Local Similarity 21.0%; Pred. No. 0.0029;
 Matches 86; Conservative 58; Mismatches 121; Indels 144; Gaps 18;

QY 35 AKREGAGDAEADP---LEP---PPQOOLGLSLRRSGROPK-RTE-RVEQFTIARRRR 87
 DB 2245 AOSSSVGPAQAQPTAQSAXPOPTQ-----PQPAQPEVQTQEVQTITVSSHVPS 2299
 QY 88 SMPVSLDSGEP---TSCPATDAETASGVSSESATSGPQSASTAVKRPASSEKVK 143
 DB 2300 AQPTAQ--SSKQVAAQSPQSNVQSPVVSQSPQIRP---STPSQLSPGQSQVQ 2355
 QY 144 GGDHDDTSDSP-----SDGLTKELQ-----NRLRRKREQEETERP- 180
 DB 2356 -----TTTSQPIPIQHTSIQIPSQSQPOQVVMKHNVAIBHLKOKKSMFTAEREE 2408
 QY 181 -----LKG1-----QSLRKKRRGEGPAETVGSSEASDVEGVLPSKQSPENDQ 223
 DB 2409 NORMIVCNQVMIKIDKIDKEQAKKRRSEVEQKSKQNAKLSALLFKHKEQLRA 2468
 QY 224 GYVSAQKDDRESKLS--GKAQADKDEPGDL-----GRPPECEG 263
 DB 2469 EILKKRALDKLOIEVQSELRDLKIKKEDLMQLAATAVAAPCPVTPAPPAPP 2528
 QY 264 YDP-----NLYC 271
 DB 2529 PSEPPPPAVQHTGLSTPTPLPAASQYKREERKSSSKKKKMMISTTSKETKKTKLYC 2588

QY 272 ICRQPHN-NRFMICDRCRWFHGDVCGVISEARGRLLENGSDYICPNC 319
 DB 2589 ICKTPVDESKEFVIGDRCQWYHGRVCGVILQSEAEI-----DEVVCPQC 2633

RESULT 9
 Q818F0
 ID Q818F0 PRELIMINARY; PRT; 542 AA.
 AC Q818F0;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Chromosome scaffold protein p85.
 OS Eukaryotes crassus.
 CC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
 CC Euplotida; Euplotidae; Moneuplotes.
 ON NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sharp S.I., Pickrell J.K., Jahn C.L.;
 RT "The identification of a novel chromosome scaffold protein that
 RT associates with rec elements undergoing en masse elimination in
 RT Euplotes crassus";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY155457; AAN75020.1; -- F0A3B5E3AA771A1 CRC64;
 SQ SEQUENCE 542 AA; 60264 MW; 60264 MW; 9785D7097C2099F9 CRC64;

Query Match 6.3%; Score 187.5; DB 5; Length 542;
 Best Local Similarity 19.4%; Pred. No. 7.1e-05;
 Matches 110; Conservative 94; Mismatches 189; Indels 175; Gaps 23;

QY 6 DPNEZAPKAIKTSKEPRKTNWGRRTTIKREGAGDAEADPLEPPPPQOOLGLSLRRSG 65
 DB 42 NPSDEE-----DINKESARTSHNKSSEKSDVEENNDDHA-----PKQEVKTRKTS 88
 QY 66 RQ-----PKRTERVEQFLTIARRRGRSRMPVSLSDSGEPTSCPATDAETASGVSSESASE- 120
 DB 89 RKTAPPKTKTKENK--VSGKTRKVKKEPDDE-----EVSEDKXVAKSSSI 135
 QY 121 -TRSGPOSASTAVKRPASSEKVKGGDDHDDTSDSDGLTKELQNLRLRRKREQEETER 179
 DB 136 LRSRLAAANAKKLSAK---KG--ENQNSDSEYEEDTPK-----SKRQAKKASK 183
 QY 180 PLKGIQSLRLKRRRGEPAETVGSSEASDVEGVL--PSKQBPENDQGVVSAQKDDRESKL 238
 DB 184 DEESSEAESE 231
 QY 239 EGKAAQDKDEPQDLGRPCEGYPDPNLYCICROPHNNRPMICCDRCRWFHGDVCGV 298
 DB 232 E-----EDDEPSDESSE 262
 QY 299 ISEARGRLLENGSDYICPNTILQVODETHSETADQEAQKWRPGDADTCTSIGTIEQ 358
 DB 263 SGCEG-----HEND-----EMEEAEGEEQDEDEMSSEAEGBE-----EE 302
 QY 359 KSSDQGIKRIEKAANPSGKKLKIPOPVIEAPGASKICIGGCHVAQPSVVCSDNCI 418
 DB 303 TVKDEGKKGKVEBEKKVAGKK-----DSPKKK-----DSPKKKKDSP 342
 QY 419 LKHAATKTLSSKEQKPK-----PRT; 443 AA.
 DB 343 GKEDVKKVSTKSKKNSPKDKKKAKEKKEKVAKKDADTKKDKTEKKKEITKDKAT 402
 QY 440 KEKMKPEKPSLPKCGAQAQIKISSVHKRPAPKEKTTVKAQVVPARSALGKEAAACE 499
 DB 403 KDKDIQKEKTSFKDSKAGTK---EHKASQSEKKKQVKS---PKDKATSKGKAE 456
 QY 500 S-----STPSWASDNYNAVPEKTAAPS 523
 DB 457 SKKKSDFDKKKDKKKISPKNGKASS 484

RESULT 10
 Q8C969
 ID Q8C969 PRELIMINARY; PRT; 443 AA.
 AC Q8C969;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN A630082K2ORIK.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=1246851;
 RA The FANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL: AK042834; BAC31377.1; --
 DR MGD; MGI:2443388; A630082K2ORIK.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR003347; TF Jm1c.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF02373; Jm1c; 1.
 DR Pfam; PF00628; PHD; 1.


```
DR SMART; SM00558; Jm1C; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Hypothetical protein.
FT NON_TER 443
SQ SEQUENCE 443 AA; 49944 MW; C543F487717C4682 CRC64;

Query Match
Best Local Similarity 22.0%; Score 187; DB 11; Length 443;
Matches 78; Conservative 44; Mismatches 118; Indels 114; Gaps 18;

QY 240 GKAAQDIKDEBGLGRPKPCBGTDPNLYICICQPHN-NRPMICCDCEBWFHGDVCG 298
Db 15 GAAAAGSAGSAPGRASAPPP-----PPVYVCQRPQYDVRNFMIECDVCKWFHSCVCG 68
QY 299 ISARGELLERNGEDYICPNTILQVQDTHSETAQOEAQKWRPGD-----ADGTDCTSIG 354
Db 69 VEEHHAVIDL-----YHCPDCAAL-----HGSLMKRRNHRHHDYTEVDDGSKPQVAG 118
QY 355 T-----IRKSSSDQG--IKGRIEKAANPSGKKLKIFQ 386
Db 119 TRAFVKELSRVFPSSADEIIVKMGHSQLTQRYLEKHGFDVPIMVPKLDDLGLG-----LALPS 174
QY 387 PVIEAPGASKICGPGCHVAQPSVYCSNDCLKHAAATMKFLSSGKEQKPKKMKMK 446
Db 175 PAFSVMQVRYVG-----GDKVIDVID-VARQADSKMTLEN-----YVKYFMN 216
QY 447 PEKPSLPKCGAAGIKISSVHKRPAPKETTIVKAVVVPARSEALGKEACSSSTPSWA 506
Db 217 PDRPKV-----LNVISL-----EFSDTKMSLVEVPDIARKL-----SWV 251
QY 507 SDENYNNAVPEKTAAPSLLYKCM-----YHLGVGLLDPSSRFWIAIPW 551
Db 252 --ENY---WPDDSVFPKPPVQKCYCLMGVQDSYTDHFIDFG-----GTSVWTHVLM 296

RESULT 11
Q8C9E0 PRELIMINARY; PRT; 473 AA.
ID Q8C9E0 AC Q8C9E0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
GN A630082K20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466653;
RA THE FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
RL EMBL; AK042327; BAC1226.1;
DR MGD; MGI:2443388; A630082K20RIK.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003347; TF Jm1C.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02373; Jm1C.1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00558; Jm1C; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
```

```
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Hypothetical protein.
FT NON_TER 473
SQ SEQUENCE 473 AA; 53409 MW; 94C94378609C7039 CRC64;

Query Match
Best Local Similarity 22.0%; Score 187; DB 11; Length 473;
Matches 78; Conservative 44; Mismatches 118; Indels 114; Gaps 18;

QY 240 GKAAQDIKDEBGLGRPKPCBGTDPNLYICICQPHN-NRPMICCDCEBWFHGDVCG 298
Db 15 GAAAAGSAGSAPGRASAPPP-----PPVYVCQRPQYDVRNFMIECDVCKWFHSCVCG 68
QY 299 ISARGELLERNGEDYICPNTILQVQDTHSETAQOEAQKWRPGD-----ADGTDCTSIG 354
Db 69 VEEHHAVIDL-----YHCPDCAAL-----HGSLMKRRNHRHHDYTEVDDGSKPQVAG 118
QY 355 T-----IRKSSSDQG--IKGRIEKAANPSGKKLKIFQ 386
Db 119 TRAFVKELSRVFPSSADEIIVKMGHSQLTQRYLEKHGFDVPIMVPKLDDLGLG-----LALPS 174
QY 387 PVIEAPGASKICGPGCHVAQPSVYCSNDCLKHAAATMKFLSSGKEQKPKKMKMK 446
Db 175 PAFSVMQVRYVG-----GDKVIDVID-VARQADSKMTLEN-----YVKYFMN 216
QY 447 PEKPSLPKCGAAGIKISSVHKRPAPKETTIVKAVVVPARSEALGKEACSSSTPSWA 506
Db 217 PDRPKV-----LNVISL-----EFSDTKMSLVEVPDIARKL-----SWV 251
QY 507 SDENYNNAVPEKTAAPSLLYKCM-----YHLGVGLLDPSSRFWIAIPW 551
Db 252 --ENY---WPDDSVFPKPPVQKCYCLMGVQDSYTDHFIDFG-----GTSVWTHVLM 296

RESULT 12
Q95YM2 PRELIMINARY; PRT; 17352 AA.
ID Q95YM2 AC Q95YM2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE I-connectin.
GN I-CON.
OS Procambarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidea; Cambaridae; Procambarus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBL J. 20:4826-4835(2001).
CC [1]-SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB055861; BAB64297.1;
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000577; FGGY kin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig c2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 49.
DR SMART; SM00018; SH3; 1.
DR SMART; SM00408; IGG2; 13.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00933; EGGY_KINASES_1; 3.
DR PROSITE; PS00835; IGLIKE; 49.
DR PROSITE; PS00002; SH3; 1.
```


KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 6.1%; Score 187; DB 5; Length 17352;
Best Local Similarity 21.3%; Pred. NO. 0.0049;
Matches 136; Conservative 81; Mismatches 206; Indels 218; Gaps 32;

QY 2 DDKGDPSENE--APKAIKPTSKPRKTWGRRTTIKREGAGD-----AEADPLEP- 50
DB DEEKLPSDEQKLKKAQKAPKKBEEIIPSRGKLPKPKDEKEIITLKPKFKVKPKPK 6267
QY 51 PPPQOQLGLSLRRSGRQKRTVERVEQFLTIARRGRSRMPVSLSDSBEPTSCPATDA 110
DB PSPKLPKPKPEPIPEPEKS-----PL-----EPYSKPEKE--V 6300
QY 111 SSGVSSESTSGSQSASTAVKPPASSEKVKGGDDHDTSDSDGLTLKELQNLRR 170
DB 6301 SDKIPELAKEVSRPEE-----PEKPEPEKL-----DESDDKPKDESETITQRP--KKRL 6348
QY 171 KREQPTERPLKGIOSRLRKRR-----EEGPAETVSGEASDTVEGVLPFS----- 215
DB 6349 QKLKAPTEKEPEIPKVTLRKTSQKVPVPEVLTETVELEHVEITPEVPEVEKRWSP 6408
QY 216 -----KQEPNDGVVSOAGKD-----DRESKL- 238
DB 6409 EYETVVPBEIPEKSPVELEKYEKYPPTKPKDDEEDKGYERKPKPKPEPEEDRK 6468
QY 239 ESKAAQDIKDBEPGD-----LGRPK-PECEGYDPNALYICICROPHNNRPMICCDRC 290
DB 6469 KKK-----LRPEEGEKKLKKPKRPKSPKPEAKPQLKPKPKPE----- 6511
QY 291 WPHGDCVGISEARGLLRNGEDYICPNCTI-----LOVODETHSETAOQEA----- 338
DB 6512 -----EKKQKVTPKPGKK-----PSKKIPDRBFVELEPFETPEPILDKKVPLEKP 6559
QY 339 ---KWRGDADGTDCTISCTIGKSSQDGKIGRIEKAANSGKKLKIFQPVIAFG-A 394
DB 6560 LKPKPEPEKPSIPEAPKPLEEKSSEEPKPKKER---PKPEK---BEEAEVPSWR 6612
QY 395 SKCIGPGCHVAQPSVYCSNDCILKHAATMKFLSSGKEQKPKPKPKM-KPKPSLP 453
DB 6613 GKRLPP---KEEKEEI-----VLKPKKPEPEKPKPKPKPKPKPKPKPKPK 6656
QY 454 K-----CGAAGIKTSSVHKRPAPKPKETTVKKAIVVPAARSALGKE 495
DB 6657 EPEKTPLEPYTKPKREKVPDGVTEPVKPED-SEKPKPE-BEIKPKKRIKPKDB----- 6709
QY 496 AACGEGSTPM-----ASDHNVAVKP---EKTAAPSPS 525
DB 6710 ---EVETPSWRGKRLPPKPEEKDEKEIITLKPKPKPKPKPKPS 6747

QY 496 AACGEGSTPM-----ASDHNVAVKP---EKTAAPSPS 525
DB 6710 ---EVETPSWRGKRLPPKPEEKDEKEIITLKPKPKPKPKPKPS 6747

RESULT 13
Q7SZX6 PRELIMINARY; PRT; 563 AA.
ID Q7SZX6
AC Q7SZX6; 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE SI:zk13a21.10 (Novel protein similar to human and mouse Cpg binding protein (CGBP)).
GN SI:zk13a21.10.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pandian R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844847; CAC30421.1; -- A084ED9C4914BB83 CRC64;
SQ SEQUENCE 563 AA; 66601 MW; A084ED9C4914BB83 CRC64;

Query Match 6.1%; Score 182; DB 13; Length 563;
Best Local Similarity 21.9%; Pred. No. 0.00018;
Matches 68; Conservative 31; Mismatches 66; Indels 146; Gaps 11;

QY 248 DEBFG-DLGRPKPECEGYDPNA-LYCICROPHNNRPMICCDRCBFWFHGDCVISE----- 301
DB 9 DQTEGLDMSKEGE-----NAPLYICTCRKSDINCWFICGDCNFWFHGHCINVTQKAK 62
QY 302 -----ARGRLLRN----- 310
DB 63 AIREWCOQCRARDPSLSIRYKKNRDKDVEPERVEKRSSTPEYKIDKRGSKVKRSARM 122
QY 311 -GEDYICFNCT-----ILOVODE----- 327
DB 123 CGE---CEPCTRTDCGCHDFCKDKKFGGPNKIQRKLRQCQVVRARKMLRVDEEFSL 179
QY 328 -----THSETAOQEB-----AKVRPGDADGTDCTISCTIGTIEKSSDQ 364
DB 180 RERKDNIMHRRYSDYDNDMDLYHYKORNASWSEDDDGLQISFVP--RKAIAVK 237
QY 365 GIKGRIEKAANSGKKLKIFQPVIE-----AFGASKCIQPGCHVAQPSVY 412
DB 238 HVKEEDKFKKESRRHKQKQKRDRLRHSRDRDGRHGSDTQCLGPNCTEPARPNSKY 297
QY 413 CSNDCILKHA 423
DB 298 CSEDCGMKLA 308

RESULT 14
Q7SZE4 PRELIMINARY; PRT; 563 AA.
ID Q7SZE4
AC Q7SZE4; 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AB; TISSUE=Body;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;
RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;
RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.P.; Bhat N.K.;
RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Heide F.;
RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;
RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;
RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Prange C.;
RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullaly S.J.;
RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;
RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Huiyk S.W.;
RA Villalón D.K.; Muzny K.C.; Sodergren E.J.; Lu X.; Gibbs R.A.;
RA Fahey J.; Helton E.; Kettman M.; Madan A.; Rodrigues S.; Sanchez A.;
RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;
RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;
RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;
RA Krzywinski M.I.; Skalska U.; Smalls D.S.; Scherch A.; Schein J.E.;
RA Jones S.J.; Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052918; AAH52918.1; --

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:51:47 ; Search time 21 Seconds
(without alignments)
2574.268 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKEDPSREAPKAIKPTS.....RSFWIAIPWACPLGVAAALC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	6.5	488	2 I46014	Cylicin II - bovin
2	189.5	6.3	706	2 A45990	junctional sarcopl
3	178	6.0	606	2 A43427	neurofilament trip
4	175	5.9	424	2 T41449	probable phd frige
5	175	5.9	5327	2 T13564	microtubule-associ
6	174.5	5.8	669	2 J05662	hepatoma-derived g
7	171.5	5.7	699	2 I38073	nucleolar phosphop
8	171.5	5.7	1110	2 I31116	NP-180 - sea lampr
9	170	5.7	510	2 T21430	hypothetical prote
10	169	5.7	729	2 S68191	triadin - human
11	169	5.7	990	2 I51618	nucleolar phosphop
12	167.5	5.6	1684	2 J00057	gravin - human
13	166	5.6	452	2 T21435	hypothetical prote
14	165.5	5.5	734	2 B42680	nucleolus-cytolas
15	165.5	5.5	1057	2 H3273	ribonuclease E pA2
16	164	5.5	1359	2 T34036	hypothetical prote
17	159.5	5.3	1087	1 QPM5H	neurofilament trip
18	157.5	5.3	405	2 T21433	hypothetical prote
19	157	5.3	1173	2 T31421	C-terminal domain-
20	156	5.2	1298	2 I54367	X-linked nuclear p
21	156	5.2	1840	2 T30250	GTL protein - mous
22	155.5	5.2	1560	2 T42727	proliferation pote
23	154.5	5.2	1588	2 T38660	probable transcrip
24	154.5	5.2	1830	2 A37981	microtubule-associ
25	153.5	5.1	1200	2 A46194	neurofilament prot
26	153.5	5.1	1233	2 T41517	serine/threonine p
27	153.5	5.1	3488	2 T34418	hypothetical prote
28	153.5	5.1	6642	2 T29757	protein UNC-89 - C
29	153	5.1	952	2 T18837	hypothetical prote

30	152.5	5.1	1641	2 I38614	helicase II - huma
31	152	5.1	763	2 T08929	hypothetical prote
32	152	5.1	1020	1 QFHUH	neurofilament trip
33	151.5	5.1	849	2 S00030	neurofilament trip
34	151.5	5.1	856	2 T16543	hypothetical prote
35	150.5	5.0	272	2 T02745	nucleic acid bindi
36	150	5.0	713	2 A28706	calpastatin, card
37	150	5.0	771	1 A33430	h-caldesmon - chic
38	150	5.0	1171	2 T13065	PIR82 protein - fr
39	149.5	5.0	644	2 S55395	neurofilament prot
40	149.5	5.0	902	2 T47966	hypothetical prote
41	149.5	5.0	1622	2 J00378	DNA (cytosine-5)-
42	149	5.0	273	2 T51145	nucleic acid bindi
43	149	5.0	810	2 T44430	protein PVI00 (imp
44	148.5	5.0	679	2 S28366	recombination repa
45	148.5	5.0	1560	2 I54361	SMCX protein - hum

ALIGNMENTS

RESULT 1

I46014
Cylicin II - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-Nov-1999
C:Accession: I46014; S52772
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.
Exp. Cell Res. 218, 174-182, 1995
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the
A:Reference number: I37271; MUID:95255491; PMID:7737358
A:Accession: I46014
A:Status: preliminary; translated from GB/EMBL/DDBV
A:Molecule type: mRNA
A:Residues: 1-488 <HES>
A:Cross-references: EMBL:Z46789; NID:G575753; PIDN:C8A86753.1; PID:G5757754

Query Match 6.5%; Score 193; DB 2; Length 488;
Best Local Similarity 20.3%; Pred. No. 0.00064;
Matches 117; Conservative 73; Mismatches 204; Indels 182; Gaps 19;

Qy	18	PTSEKPKTWGPRRTIAKREGAGDAEADPLEPPPPQQLGLSLRRSGRQPKRTERVEQF	77
Db	19	PVSELSKSN-----QHFALVFPKPPKPKR-----	46
Qy	78	LTIARRRRGRSMFVSLSDSGEPTSCPATDAE-----TASEGSVESAS	119
Db	47	-----RSKPSLLQEN-----TSPKYDAEKLGRDKQPLMWHRSELRISRPSVYLA	94
Qy	120	ETR-----SGFQASATAVKERPASSKVKGGDDHDDTSDSDGLTKELQNLRRKKEQ	175
Db	95	RSRHPQKETPPSOEDAKQAAPSPKVKSKEDKXSDSEASISVSKPKRKLKAKEEK	154
Qy	176	PTER-----PLKGIQS-----RLRKKRREGPAFTVGSASDT	208
Db	155	PDEKDLKKERKSKGKESATSEDEKAGAKGKDKGSKKGTPTSDS--GSEKGA	213
Qy	209	VGVLPSKQPE-----NDQGVVSQAGKDRSK-----LEGKAAQDIQEE	250
Db	214	KDSKSKGSKGKESATSEGEKGDADKDKKKGKSGKKGESATSEGEKGDADKDK	273
Qy	251	PGDLGRK-----PECRGYDNLALYCICRQPHNNRPMICCDRCBFWHGDVCSHARG	305
Db	274	KGKSGKKGESATSEGEKGDADK-----DDKKGK	304
Qy	306	LLERNGBDYICPNCTILQVQDTHSETADQEAQRPGDADGTCTCTGTYEQKSSDQ	365
Db	305	KGSKGKE-----SATESGEKGDADKDKKKGKSGKKG-----KESATSEGE	347
Qy	366	IKGRIEKAAPSGKKLKIQQPIEAPGASKIGCGCHVAQPSVYCSNDILKHAAT	425
Db	348	EKGDAKK-DDKKGKK-----GSKK--GKESDSKAEKGDGAKKD--DKDKKGS	391

[illegible]

RESULT 2

A45990
 Junctional sarcoplasmic reticulum glycoprotein triadin - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 11-Jan-2000
 C:Accession: A45990; S45704
 R:Knudson, C.M.; Stang, K.K.; Moomaw, C.R.; Slaughter, C.A.; Campbell, K.P.
 J. Biol. Chem. 268, 12646-12654, 1993
 A:Title: Primary structure and topological analysis of a skeletal muscle-sp
 A:Reference number: A45990; UID:93286104; PMID:7685347

A;Accession: A45990
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-706 <GNU>
A;Cross-references: NID:g347850; PIDN:AAA31488.1; PID:g347851
A;Experimental source: triads, muscle
A;Note: sequence extracted from NCBI backbone (NCBIN:133603, NCBIPI:133604)
P;Peng, M.; Pan, H.; Kirely, T.L.; Caswell, A.H.; Schwartz, A.
FEBS Lett. 348, 17-20, 1994
A;Title: Structural diversity of triadin in skeletal muscle and evidence of its existend
A;Reference number: S45704; MUID:94289946; PMID:8026576
A;Accession: S45704
A;Molecule type: mRNA
A;Residues: 401-620 <PEN>
C;Superfamily: histone H1
C;Keywords: glycoprotein

Query Match 6.3%; Score 189.5; DB 2; Length 706;
Best Local Similarity 22.9%; Pred. NO. 0.0015;
Matches 99; Conservative 62; Mismatches 177; Indels 95; Gaps 16;

Qy	145	GDHDDTSDSDGLTJKE--LONELPKEKEQEPTPELKGIOSL-----RKKRBERGAE	199
Db	118	GDDEDEGEEDTAKGEIEEPLKKEIOHKKEIKEQEPERKPIPTKVHKEKEKEKVKE	177
Qy	200	TVGSEASDTVSGVLPKQEPENDEGVVSQAGKHDDRESKLEGTAAQDIIDBERPDLGR---	256
Db	178	KEPKGKAHAKLEKKEKEKET--KTVTKESKARTKEKIEENTKEVAGVQKEVKYQTVTA	236
Qy	257	-----PKP-EEGVDDNALYICIQPHNNRPMICCDRCBEWHFHCQVCGISEARGELL	307
Db	237	KAEVQKTPPKKEKESKETAA---VSKQEQDQYAFRYMIDIFYHGD-----L	282
Qy	308	ENGSDYICPCTTILQWDETHS--ETADQOEAKWREGDADGTDCTSIGTIEQKSESDQGI	366
Db	283	KFGQSPALPPSPTEQASRPTPALPTPEKEGSKKAE-----KKVTETKGAEKEDA	336
Qy	367	KGRLEKANPSGKKKLKIFOPVIEAPGASKCIPGCHVAQSPDSVYCSNDGILKHAATM	426
Db	337	KKKSKKETDIIMKKK-----EKGSPDTKPGTVKVTI-----QAATK	373

427	QY	---KFLSSGKECP---KSKCKMKCKEKSLSLPCGNA-----QAGIKI-----	463
374	DB	KDEKEDSKAKPAKBAECPKGGCKGCKEKEEPAKSTKKEHAAPSKQAKLKERKEVS	433
464	QY	SSVHKRPAPEKKTETTVKAVVVPARSALGKEAAC-----BSSTPSWASDHNYNA	513
434	DB	AASTTKAVPAKKBEKTKNTVEQETRKEKPKGISVLNDELTKETKEVKVPASLKKGSET	493
514	QY	VKPEKTAAPSPSL	526
494	DB	KDEKTSKEPEOI	506

RESULT 4

RAS001: 1
 T1449
 Probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 C:Accession: T11449
 R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z11994
 A:Accession: T11449
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-424 <RIE>

RESULT 3

A43427 neurofilament triplet H1 protein - rabbit (fragment)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 10-Dec-1999
 C:Accession: A43427
 J:Roppet, D.R.; Beasley, L.L.; Willard, M.B.
 J Biol. Chem. 257, 17154-17361, 1992
 A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polypeptide
 A:Reference number: A43427; PMID:92381055; PMID:1512270
 A:Accession: A43427
 A:Molecule type: DNA
 A:Residues: 1-606 <SOP>
 A:Cross-references: GB:M94315; NID:gl64990; PIDN:AAAS7152.1; PID:5601930
 A:Experimental source: brain
 A:Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBI:P:112011)
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.0%; Score 178; DB 2; Length 606;
Best Local Similarity 19.9%; Pred. No. 0.0057;
Matches 100; Conservative 67; Mismatches 211; Indels 124; Gaps 17;

Qy	35	AKRGACDAADADPLPPPPPOOGLSLRRGRGPKRTVEQFLTTARRGRRMVSYLE	94
Db	28	AKSEEGGEEK---EAKSPFGGAASPEENKGAFAEK-----SPK-E	66
Qy	95	DSGETSCPADTAETAGSGVSASERTSRCSQASATVKERPASSKVGGDHDDTSDTS	154
Db	67	EAKSP-----AAKAPAEAKSPAEAKS-PEKAKSPVKEAKSPKAEKAKSPA	118

Qy 155 DSDGLTKLQLNRLLRRKREQEPETPLKGIQSRLLKRRREEGPABTVGSE-ASDTVEGVL 213
Db 119 EA-----KSPKAKSPABAKSPKAKSPVTVBEAKSPETAKSPABAKSPKAKSPABAKS 172

[illegible]

Db	173	PEKAKSPVKEAKSPERAKSPVKEAKSPAEAKSPERAKSPVKEAKSPERAKSP-----	227
Qy	265	DPNALYCTCRQPHNNRFMICCDRCCEWTHFGDCVGISSEAGRLLRRNGEDYICPNCTILQV	324

```
Db      228 -----:|:|:|-----EA 238  
                ARAKSPVKB  
Cv      326 ANPTTSEPTANQDPAVWZBDNADNDNDGCMSTCTEYVSSSPPQCTGVGBTPVANDSCVWTWT 384
```

QY 325 QDTHSETADQEARWFGDADGDC*SIGTIRQSSDEQGIKRIEKAANPSGAKKRLI 388

DB 239 KSPEKAKSPKEBAK-SPAEXS-----PEKAKSPKAKSPVEVKSPEAKSPEKA 288

QY 385 PQPV-IEAPGASKICIGPCCHVAOPDSVVCNSDILKHAAATMKFLSSGKQKXP-KKK 442

D6 289 KSPVKEAKSPKAKSPVKEAKSPEK---AKSPVKEAKSPEKAKSPVKEAKSPEKAK 345

Qy 443 MOKPKPSPSLPKCGAQAGIKTSSVHKGPAPEKGGTTTKAAVVVPARSEALGKEAACESST 502

Db	346	SPVKEBAKSPKAKSPVKBEAKSPKAKSPKBEAKSPKAKSPVKE	---EAKS 402
QY	503	PSWASDENYNAVKPEKTAAPSP	524

Db 403 PE-----KAKSPVTEAKSP 417

RESULT 4

RAS001: 1
 T1449
 Probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 C:Accession: T11449
 R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z11994
 A:Accession: T11449
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-424 <R1E>

Db 323 RDEERRLEARRRREQEELRLRE-QEREKERRKE-RAERGGSS-----SCE 369
Qy 218 ERENQGVVSQAGKDDRESKLGKAAQDIKDEP-GDLGR-----PKPECEGYDPNLY 270
Db 370 ELEDEEPV-----KKRSKARGTTPSSSDPEPEGLGKEKKLAKKSQLPSSS-----419
Qy 271 CFCROPHNNRFMICCDRCBFWHGDVCGISRGARGLLERNGBDYICPNCTILOVDETHS 330
Db 420 --ARKP-----GQKEKRG-----431
Qy 331 ETADQOEAKWRPGDADGTCTSIGTIEOKSSSDQIGKRIEKAANPSGKKLK-----I 384
Db 432 ---PDEKPARPVKVERT-----KKRSGLSLERKGEKKPSVEERLQKLHSEIK 479
Qy 385 FQPVTEAPGASKCIGP---GQCHVA-----QPDGV-----YCSNDCILKHA--423
Db 480 FALKVDNPDVVKLSALEBLGTQVTSQILQKNTDVVATLKIRRYKAKNDVMAKAARVY 539
Qy 424 -----ATWKFLSSGKEQKPKPKKWKMKPEKPSLPKCGAAGIKISSVH-----467
Db 540 TRLKSRVLGPKVEALQKVNKAGAKERADNKKLEBQF-----GEOAPRELAEDPEST 591
Qy 468 KRPAPKKTIVKCAVWPARGAALGKZAAECSSSTPSWASDH-----NNAVKP-----516
Db 592 DRAPVNGEATSGKEMEDRAQEDQDS--EDGPRGSSSEELHSDPSRNSDPKPGNER 649
Qy 517 ----EKTAAPSPS 525
Db 650 QDHERTRLASES 661

RESULT 7

I38073
nucleolar phosphoprotein pl30 - human
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 24-Sep-1999
C:Accession: I38073; S52292
R:Pat. C.Y.; Chen, H.K.; Shen, H.L.; Yeh, N.H.
J. Cell Sci. 108, 1911-1920, 1995
A:Title: Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein P
A:Reference number: I38073; PMID:95386590; PMID:7657714
A:Accession: I38073
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-699 <RES>
A:Cross-references: EMBL:Z34289; NID:G663007; PIDN:CA84063.1; PID:G663008
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.7%; Score 171.5; DB 2; Length 699;
Best Local Similarity 19.5%; Pred. No. 0.016;
Matches 108; Conservative 84; Mismatches 218; Indels 145; Gaps 19;
Qy 2 DDKGDPNNEAPKAIKPTSKFRKTGWFRRTIAKREGAGADABAD-PLPPPPQOQLGSL 60
Db 136 DBEDQKQPVQKGVPOAKAKAP-----PKAKSDSDSDSSDEPPKQ-----183
Qy 61 LRSGROPKKTERTVEQPLTARRGRSMPSVLEDSGEPTSCPATDAFTASEGSVESASE 120
Db 184 -----KPKIT-----PVTYKATKAPKPKPABAAPKPIANGKAASSSS 219
Qy 121 TRGPOSASTAVKRPASSK-----VKGDDHDDTSDSDGLTKELQNR 167
Db 220 SSSSSSSSDSBEKAANATPKTVPKQVAKVPAATPTTKSSSSSDSDSDEBEQK 279
Qy 168 LRKROEP-----TERPLKGIOSRLRKRRREGPAETVGSASDVTGVLPSKQEP 219
Db 280 KPMNKGPSYAPPSAPPKSLGTQPPKAVKQPVESSEDSDSDSSEBEKFP 339
Qy 220 ENDQGVVSQAGKDDRESKLGKAAQDIKDEP-GDLGRPKPECEGYDPNLY-----269
Db 340 PT-KAVVSKATTKPPPAKKAASSSSSDSDSDEDEAPSKPAGTITNSNKPAPVTTKSP 398

Qy 270 ----YCIQROPHNNRFMICCDRCBFWHGDVCGISRGARGLLERNGBDYICPNCTILOVQ 325
Db 399 AVKEAAPKQP-----VGQGG-----KLTTRKAD-----SSS 425
Qy 326 DETHSETADQOEAK-----WRPGDADGTCTSI-----CTIEOKSSSDQIGKRIEKA 373
Db 426 SEESSSSSEETTKQWATYTP-KATAKAALSLEPAKQAPQGSRDSSSDSDSSSEBEK 484
Qy 374 ANPSGKKLKIQFVIEAPGASKCIGCGCHVAQPDVSYCSNDCILKHAATMFLSGSK 433
Db 485 TSKAVKK-----KPKVAGGAAP-----SKPASA-----KKGRAESSNSSSD 523
Qy 434 BQKPKPKKMKK-PKPSLPKCGAAGIKLSSVHKRPAPKPKKTIVKCAVWPAPRSAL 492
Db 524 DSSSEEBKLGKSGSPRQAPFANGTALTQNGKAAKNSSEBEKKAARVAVVSKSGSL 583
Qy 493 CK-----EAACSSSTP 503
Db 584 KKRQNEAAKEATP 598

RESULT 8

I51116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I51116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A:Reference number: I51116; PMID:7770000
A:Accession: I51116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1110 <JAC>
A:Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632549
C:Superfamily: neurofilament triplet H protein

Query Match 5.7%; Score 171.5; DB 2; Length 1110;
Best Local Similarity 20.0%; Pred. No. 0.025;
Matches 111; Conservative 86; Mismatches 233; Indels 125; Gaps 23;

Qy 2 DDKGDPNNEAPKAIKPTSKFRKTGWFRRTIAKREGAGADABADPLPPPPQOQLGSL 61
Db 486 DEEEEEEHEHEHEHEHEHEHEEDRG-----RKGEAEEAEEAEEVEKEAEEAEV 538
Qy 62 RESGROPKKTERTVEQPLTARRGRSMPSVLEDSGEPTSCPATDAFTASEGSVESASE 121
Db 539 EEAEEETEAEAAE-----EHEAEGHEEAEGEAEAEAEVEEBAIEKAEAA 587
Qy 122 RSGPOSASTAVKRPASSKVKGGDDHDDTSDSDGLTKELQNR-----REK 171
Db 588 -----EAKAEVEEEAEE-----EEBEAEEVEAETKEEVEAEEAEEVEEAGEAEEA 638
Qy 172 REQPTERPLAGIQSLRLKRRREGPAETVGSASDVTGVLPSKQEPNDQGVVSQAK 231
Db 639 EEEAEEHEVTSKKATQAEVEHEEAEEAEEAEEA-----EAGEEDVEAESKE 691
Qy 232 DDRESKLEKAAQD-----TKDEPGLGRPKPECEGYDPNLYCIQROPHNNRPMTC 285
Db 692 EEEEDSKAADAEDEAEEVEKEEVEVTKSDAEAEAEAEAE-----734
Qy 286 DRCEWFHGDVCGISRGARGLLERNGBDYICPNCTILOVDETHSETADQOEAKWRPGDA 345
Db 735 -KSEE-----EAAEAKDEAEEAEE-----EAVEETEAT-EEAEEKASDDE 777
Qy 346 DGTDTCTSIGTIEOKSS-----DQGIKGRIEKAANPSGKKLKIQFVIE-----APASKI 398
Db 778 KEKE-----EVKSEAPVAPKAKPEPKAA-PKKAPAKAVESPTSEPEDEPAEVVE 829
Qy 399 GFGCCHVAQPDVSYCSNDCILKHAATMFLSSGKEQKPK--PKKMKMKPKPSLPKCG 456

Db 830 KKGKAEAPKPKA-----KPAAKKEAKPVKEKEEPEESTER---EPKPPAAK-P 876
QY 457 AQAGIKISSVHKPAPKEKTTVKKAVVVPARSE-ALGKBAACESSTPWSADHNTNAYK 515
Db 877 AKAPAK-----PKPAP-KAEAEKZPAPKPAQAKPAPAAAEHEDEKEDDEEEVEEVK 930
QY 516 PEXT-----AAPSPS 525
Db 931 PEDAKPVKSPAPAP 944

RESULT 9
T21430
hypoetical protein F26H11.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21430
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21430
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-510 <MIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04195.1; GSPDB:GN000020; CESP:F26H11.3b
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3b
A:Map position: 2
A:Introns: 116/1; 147/3; 288/2; 348/2; 392/1
C:Superfamily: bromodomain homology
P:371-426/Domain: bromodomain homology <BRO>

Query Match 5.7%; Score 170; DB 2; Length 510;
Best Local Similarity 20.4%; Pred. No. 0.014;
Matches 67; Conservative 38; Mismatches 106; Indels 118; Gaps 11;
QY 101 SCRATDAFTASGVSASATSGPQASST-----AVKRPASSKVKGGDDHDDTSD 153
Db 31 SCPVQDQADASLEISDPKFAVLKPFBSAQDWSFGYLLAEQPGTS----- 76
QY 154 SDSGLTLKEL-----QNLKREGEOPTEPLKGIQSLRKKRREGSPAE- 199
Db 77 SDKTTTPIKTIWFKPVPIGKPPRRRCADREISLAAK---PKAEVKEVINPADI 133
QY 200 TVGSEADTVGVLPL-----SKQEPENDQGVVSOAGKDDRESKLEK 241
Db 134 TLGGDTYDYVKEQKPTESLNTVSERRRTSANTLSKSDRDKPSPQSTAPKSKERTSEP 193
QY 242 AAQIKDEBPGD-----L 254
Db 194 PASHVAFHTPGSATPHDNLNLSIEHTCOKIPDASKLYIQELCARWTHGCVGVABQTIL 253
QY 255 GRPKPECEG-----YDPNLYCICRQPHNN-RFMICCDRCBFWHGCVCVIGISEARGR 305
Db 254 GLEHWSCEECIEHEQERVKQDQALYCVQKPYDDIKFYVGCDCSQGWTHPECVGTTAAZ- 312
QY 306 LLERNGBDYICPNTILQVQDTHSETAD 334
Db 313 ---EQAADYNGFACT--REAEFGYSEASD 336

RESULT 10
S68191
triadin - human
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 11-Jan-2000
C:Accession: S68191
R:Tsake, N.L.; Eyre, H.J.; O'Brien, R.O.; Sutherland, G.R.; Denborough, M.A.; Foster, P.
Eur. J. Biochem. 233, 258-265, 1995
A:Title: Molecular cloning of the cDNA encoding human skeletal muscle triadin and its lc
A:Reference number: S68191; MUID:96061957; PMID:7588753
A:Accession: S68191

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-729 <TAS>
A:Cross-references: EMBL:U18985; NID:G982222; PIDN:AAA75315.1; PID:G982223
C:Genetics:
A:Map position: 6q22-6q23
C:Superfamily: histone H1
Query Match 5.7%; Score 169; DB 2; Length 729;
Best Local Similarity 19.3%; Pred. No. 0.023;
Matches 101; Conservative 83; Mismatches 168; Indels 172; Gaps 22;
QY 90 FVSLSDSGPTSCPATDAFTASGVSASA-----SET-----RSGPOSASTAVKE 134
Db 293 PLPTEQASRPT--PASPALREKEGKKAEKKVTSSTKKKKKEDIKKKSEKETAIDVEKK 350
QY 135 RPASSEKVKGG-----DHDHDSDSGLTLKELQNLRLRRRQ--EPTERP 180
Db 351 EFGKASRQGVTKIAQAQAAKDKKESKTKKPAEVEQPGKQKQKKEGHEVPAKSP 410
QY 181 LK--GIQSLRKRREGEPAETVGSSEADTVRGVLPSPKQEPENDQGVVSOAGKDDRESKL 238
Db 411 KKEHSPVSDQVKANTERAKEBIGAVSS---KKAVPKKEEKTTKTVEQIRKEK----- 462
QY 239 EGKAAQDIIEE-----PGDLCRPKECEGYDPNLYCICRQPHNNRPMICDRCR 289
Db 463 SGTSSILDKKEPIKGEKPKVPASLKEKEPETK----- 495
QY 290 ENPHGDCVGISEARGRLRNGEDYICPNTILQVQDTHSETADQOQAKWRFGDADGTD 349
Db 496 -----KDEKMSKAG---KEVKPKPPQLQGG- 517
QY 350 CTSIGTIEQKSHDQIKGRIEKAANPSGKKLKIPOPIEAPGASKCIGPGGCHVAQPD 409
Db 518 -----KKEKPEPQ-----IKBAKPAISEKVIHKQDIVKP--BKTVSHG-----KPE 558
QY 410 SVYCSNDILKHAATMKFLSSGKQKPKPKKKK-----MKDEKPSLPCG-----A 457
Db 559 -----EKVLQ-----VKAVTIETAKPPTKKAHREPPPSIKTDKPKPTFGTSEVT 608
QY 458 QAGIKTSVHKPAPKETTVMKAVVVPARSEALKEACESSTPSWASD--HNTNAYK 515
Db 609 ESKKKTEISEKESKADMKHLREKVKSTRKESLQLNHNTVTKAEKPAFVSKQVEDVPASK 668
QY 516 PEK-----TAPSPSLLYKCMVHLGVGLDPSSEFWIAIPW 551
Db 669 KAKEGTEDVSPTKQKSPISFPQCVY-----LDGYNGYGFQPPF 706

RESULT 11
I51618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I51618; S57757
R:Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: I51618; MUID:96019267; PMID:7593294
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: mRNA
A:Residues: 1-990 <CAL>
A:Cross-references: EMBL:X88927; NID:G895920; PIDN:CRA61368.1; PID:G895921
C:Genetics:
A:Gene: xNopp180
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.7%; Score 169; DB 2; Length 990;
Best Local Similarity 19.0%; Pred. No. 0.031;
Matches 102; Conservative 54; Mismatches 167; Indels 192; Gaps 16;


```

QY 7 PSNEAPKAIKPTSKFRKTWGRFTTIAKRG-----AGDAEADPLEPPPPQOQLGLSLR 62
Db 484 PTNSKA-----TPTSKTKTAKPGTPTSAKKDSSSDSSDSEKTKTP-----528
QY 63 RSGRPQKTERVEQFLTIARRRGSRMPVSLSDSEPTSCPATDAETASEGVSASSTR 122
Db 529 -AKRAAKTTPAKPAKTTAKPAKATTPGKQVPTKESSSDSDSSDSEDE 587
QY 123 SGPOQASTAVK-----ERPASSKVGDDHDDTSDSDGLTLKELQNLRLKRRQOE 175
Db 588 K-KSSAKPAVKTTPGKATSKPVVASKPVPAKKAASSSDSDSS-----EETT 633
QY 176 PTERLKGISRLRKKRREEGPAETVGSRASDTVEGVLPSKOEPPNDQGVVSQAGKDRE 235
Db 634 KTKKPL-----TKLSPAVKILFPKKAESSDSDSSDSEKTKTP 672
QY 236 SKLEGKAAQDIKIDBERPGDILGRPKCEGYDPNALYICICRQPHNNRPMICCDRCBFWFHGD 295
Db 673 AKPPAKSATPVNTKAPAKQKASKASC-----698
QY 296 CVGISEARGILLERNGEDYICPNCITLQVDETHSTADQOEAKWRPGDADGTDCTSIGT 355
Db 699 -----SDSD-----702
QY 356 IEQKSSDQCIKGRIEKAAANPSGKKLKIQFVIRAPGASKICIGPCCHVAQPDVSYCSN 415
Db 703 ---SSSEEG-----KSKPTGK-----SPAARATAPPK-KNPVAVNKKDPSSSSSD 746
QY 416 DCILKHAAMTKPLSGKQKQPKPE-----KMKQKPEKSLPKCGAAGIKLSSVHK 468
Db 747 -----SSGDDKQPKQAAAAKDVQGAKAAPT-PKKAASSSDSSDSE 791
QY 469 RPAPEKKTIV--KKAVVVVPARSEALGKEAACESS 501
Db 792 DVSKAKTNTVNSKSPVTPKAVPAKKESSSESS 826

RESULT 12
JW0057
gravin - human
C:Species: Homo sapiens (man)
C>Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
R:Accession: JW0057
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial
A:Reference number: JW0057; MUID:98269042; PMID:9604001
A:Accession: JW0057
A:Molecule type: mRNA
A:Residues: 1-1684 <SAT>
A:Cross-references: DDBJ:AB003476; NID:G2081606; PIDN:BAAL19927.1; PID:d1020716; PID:g208
C:Comment: This protein regulates cell growth.
F:433-439/Region: nuclear regulation signal
F:522-527/Region: nuclear location signal
F:591-596/Region: nuclear location signal
F:671-676/Region: nuclear location signal

Query Match 5.6%; Score 167.5; DB 2; Length 1684;
Best Local Similarity 19.8%; Pred. No. 0.066;
Matches 136; Conservative 73; Mismatches 218; Indels 261; Gaps 29;

QY 3 DKGDPSNEAP--KAIKPTS-----KSPRTW-GFRRTT-----33
Db 169 EGEKQKQKPSKSAESPTSPVTSSTGTFKPTQGWACWRKTSFRKPKEDVEASEK 228
QY 34 -----IAKREGAGDAEADPLEPP-----PQOQLG 58
Db 229 KIQEPKPKVDTEGKAVASEKLTASEQAHPQEPASAEHPRLSAEYKVELFSEQV- 287
QY 59 LSEPSRQ-----PKETERVEQFLTIARRRGSRMPVSLSDSEPTSCPATDA-ET 109
Db 288 -----SGSQGPESEKPAFLATEVFEKIEVHQEVEVAVHVSTVE---ERTBEQKTEVEET 340

```

```

QY 110 ASEGSVESASSTRSGPOSASTAVKRPASSEKVKVGGDDHDDTSDSDSGDLTLKE-----163
Db 341 AGSVPAELVEMDAEPQAEPA-XELVKLKETCVSGEDPTQAGADLSPDEKVLSPKPPGAV 399
QY 164 -----LQNLRLRKREBP-----TERPLKGIQSLRLKRRREG-----PAETVGS 203
Db 400 SEVEMLSQBERMKVQSGPLKCLFTSTGLKLSCKKQKGRGGGDEESGHTQVPADSPDS 459
QY 204 EASPTVEGVLPSKQBP-----NDQGVVSQAGKDDRESKLEG-----240
Db 460 QEEQKSGSSASSPREPEIITCLEKGLABVQODSEAEAGATSDGKKREGVTPWASPKXW 519
QY 241 -----KAAQDIKIDBERPGDILGRPKCEGYDPNALYICICRQPHNNRPMICCDRCBFWFHGD 270
Db 520 TPCKRVRRPSSDKEDLDKVKSATLSSTESTASEMQEEMKGSVEBPKE-----569
QY 271 CICEQPHNNRPMICCDRCBFWFHGDVCSISARGELLERNGEDYICPNCITLQVDETHS 330
Db 570 ---EPKPK-----VDTSVSWELICVSSSKRA-----RRGSS-----SDEEG 605
QY 331 ETA---DQOEAKWRPGDAD-GTDCTSIGTIEQKSSBQDQIKGRIEKAAANPSGKKLKIQ 386
Db 606 PKANGGDHOKADEAGKDKETGTDGILAGS--QEHDPGQ--SSSPQAGSPT-----653
QY 387 PVIEAPGASKICIGPCCHVAQPDVSYCSNDCILKHAAMTKPLSGKQKQPKPEKMKW 446
Db 654 ---EGEGVS-----TWESFRLVT-----PRKSKSK 677
QY 447 PEKSLPKCGAAGIKLSSVHKRPAPEKKTIVKAVVVPARSEALGK--BAACESSTPS 504
Db 678 LEEKSEDSI-AGSGVEHSTPDTPEKQESWVSIKKFTIPGRKKRPDQGRQEQAPVEDAGPT 736
QY 505 WASDIN-----YNAVKPEKTA 521
Db 737 GANEDSDVPVAVPLSEYDAVEREKMEA 764

RESULT 13
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <WIL>
A:Cross-references: EXBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CESP:F26H11.3c
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
F:313-368/Domain: bromodomain homology <BRO>

Query Match 5.6%; Score 166; DB 2; Length 452;
Best Local Similarity 21.0%; Pred. No. 0.02;
Matches 55; Conservative 31; Mismatches 94; Indels 94; Gaps 8;

QY 161 LKELQNLRLRKREQEPTE-----RPLKGIQSLRLKRRREGPAE-TVGSEAS 206
Db 23 LKKEINKRTRMENEAEQCGLLTPWRKARARPHAAKPAEYKKEVINPADITLGGDTY 82
QY 207 DTVEGVLP-----SKQEPNDQGVVSQAGKDDRESKLEGKAAQDIK 248
Db 83 DYVKEQKPTSIATNVRRRTTSANLSSEDDRDKPESOSTAPKSKERTSPASHVAF 142
QY 249 EBPQD-----LGRPKPEC 261

```

Db 143 HTPGSATPHDINLSIEHCTCQKIFDASKLYIQCBLCARWYHGDGCVGVAEQITLGLHWSC 202
Qy 262 EG-----YDPNLYCICQPHNN-REFMCCDRCREWPHGDCVIGSEARGLLRNGE 312
Db 203 EECIEBERVKDQALYVCVCQKFPDDTKFYVGCDSQCGWFAPECVGTTRAEA-----EQAA 258
Qy 313 DYICPNCITLQVQDTHSETAD 334
Db 259 DYNCPACT--REARGYSEASD 278

RESULT 14
B42680
nucleolus-cytoplasm shuttle phosphoprotein - rat
N:Alternate names: Nopp140 protein B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 17-Nov-2000
C:Accession: B42680; A42680; S27889; S30510; S30511
R:Meier, U.; Blobel, G.
Cell 70, 127-138, 1992

A:Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542; PMID:1623516
A:Accession: B42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <ME1>
A:Cross-references: GB:M94288
A:Experimental source: clone pTM6
A:Accession: A42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'Q', 181-734 <ME3>
A:Cross-references: GB:M94287; NID:9205749; PIDN:AAA41718.1; PID:G205750
R:Meier, U.; Blobel, G.
submitted to the EMBL Data Library, May 1992
A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: S27889
A:Accession: S27890
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:G205751; PIDN:AAA41719.1; PID:G205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, 'Q', 181-734 <ME12>
A:Cross-references: EMBL:M94287; NID:G205749; PIDN:AAA41718.1; PID:G205750
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.5%; Score 165.5; DB 2; Length 734;
Best Local Similarity 21.7%; Pred. No. 0.036;
Matches 121; Conservative 75; Mismatches 229; Indels 133; Gaps 25;

Qy 3 DKGDPSNEEAPKAIKPTSKBPKTWGFRRTTIAXR-----EGAGDAEADPLRPPPPQQLG 58
Db 117 DSSSDSSEEDKAQVPTQK-----AAPAKGASLPQAGKAAKASSSSSSE--- 164

Qy 59 LSLRRSGQPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASEGSVESA 118
Db 165 -SSREBEKDKKKVPQKAVKPAKAVR--PPPKKASSESESDDSSSEDEAPQTKPKAA 221

Qy 119 SETRSGPOSASTAVKERPASEKV-----KGGDDHDDTSDSDSGLTLELQNLRLRK 171
Db 222 ATAAKAPTKAQTAPKAPGPPAKAQPCKAANGKAGSSSSSSSSSSSSDD-----S 269

Qy 172 REOSPTERPLKIGISRLKRRKEEPAETVGSASDITVEGVLPSKQEPENDQGVVQAGK 231
Db 270 EEKKAAPLK-----KTPAKQGVAKAPVKVTAAPTQKSSSD-----SSSEE 314

Qy 232 DDRSKLEKGAQDIKDEPGDLGRPKPECEGYDNPALYICRQPHNNRPMICDRCSEW 291
Db 315 EEEQKKMKKAGPYSSVPPPSVLSKKSVAQSPKKA-AAQTQPADS-----SADSSSE- 368

Qy 292 FHGDCVIGISEARGLLRNGEDYICFNCTILQVQDTHSETADQQAQKWRPGDAGTDCT 351

Db 369 -----SDSSSE-----HEKTKPAKTV-----SKTPAKPAVKKKARSSSDSSDSD-- 409
Qy 352 SIGTIEKSESDQIGRIEKAANPSGKKLKFQV-VIEAPGASKCI-----GFGCC 403
Db 410 -----SSED-----EAPAKPVSAKSPKSPKPAVTPKPPAKAVATPKQAPAGSG-- 452
Qy 404 HVAQPSVYCSNDCILKHAATAWKFLSSGKEQPK-----PKEKMKMKPEKPSLP-KCGA 457
Db 453 --QKQSRKAD-----SSSSEESSSESEATKSKVTTTPKARVTKA-AAPSLPAKQAP 502
Qy 458 QAGIKUSSVHKRPAPKPKETTIVKAVVVPARSALG-----KEAACESSTPWSAD 508
Db 503 RAGDSSDSSSESSSEKKTTPKP-----PAKKAAGAATVPKPTPVKKAARESSSSSSSE 559
Qy 509 HN--YNAVPEKPTAAPSP 524
Db 560 DSSEEEKKPKSKATPKP 577

RESULT 15
H83273
ribonuclease E PA2976 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 31-Dec-2000
C:Accession: H83273
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83273
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1057 <STO>
A:Cross-references: GB:AE004723; GB:AE004091; NID:9949067; PIDN:AAG06364.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: rne; PA2976
C:Superfamily: ribonuclease E

Query Match 5.5%; Score 165.5; DB 2; Length 1057;
Best Local Similarity 21.5%; Pred. No. 0.053;
Matches 126; Conservative 76; Mismatches 228; Indels 155; Gaps 26;

Qy 7 PSNEEAPKAIKPTSKBPKTWGFRRTTIAXRSEGAGDAEADPLE--PPPPQQLGLSLERS 64
Db 535 PCHTEAPVEAPKMPPEPSLPQGLVKSILVGLFAGKQPAKPAETSKPAERQTRQDSRN 594

Qy 65 GRQPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASEGSVESASETRSG 124
Db 595 GRQNNRRDGRD---GNRRDESKPKPEERAERQPRE-----ERAERPNREESERRR 644

Qy 125 POSASTAVKERPASSEKVKGGDDHDDTSDSDGLTLKBLQNLRLRKREDEPTER--P 180
Db 645 ERA-----ERPAREER-----QPREGRERAERTFREERQ 675

Qy 181 LKGIQSR-LRKRRREGPAETVGSASDITVEGVLPSKQEPENDQGVVQAGKDRSK-- 237
Db 676 REGREGREESERRRERAEERPAERERQPREGRERAERPAEREE---RQPREDRQARDA 731

Qy 238 --LEKKA-----AODIKDEEPCDLGRPKPECEGYDNPALYICRQPHNNRPMICDRC 288
Db 732 AALAEALPNDESLEQDEQDDTQGE--RPRRRSGQ-----RRRSNR-----ERQ 775

Qy 289 EEFPHGDCVIGISEARGLLRNGEDYICFNCTI-----LQVQDE-----THS 330
Db 776 RE-VSGLEEG-SEAT-----DNAAAPLNTVAATAAGAAAGIASEAVANVEQAPATTS 825

Qy 331 ETADQQAQKWRPGDAGTDCTSTGTIEKSESDQIGK-----IEKAAPNSGKK 381
Db 826 EAASETTA-----SDETDASTSEAVETQADSEANTGETADIEAPVTVSVVRDEADQST 879

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 08:47:21 ; Search time 59 seconds
(without alignments)
2691.385 Million cell updates/sec

Title: US-09-787-016A-3
Perfect score: 2989
Sequence: 1 MDDKGDPSNEAPKAIKPTS.....RSFWIAIPWACPLGVAMLC 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq19808:*
2: Geneseq19908:*
3: Geneseq20008:*
4: Geneseq20018:*
5: Geneseq20028:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2989	100.0	562	3 AAY67579	Aay67579 Human dea
2	2981	99.7	562	4 AAB93638	Aab93638 Human pro
3	2431	81.3	1191	4 AAM80219	Aam80219 Human pro
4	2311	77.3	647	4 AAM79235	Aam79235 Human pro
5	2305.5	77.1	775	4 ABG22389	Abg22389 Novel hum
6	2108	70.5	614	3 AAY67580	Aay67580 Murine de
7	1551	51.9	330	4 ABG22388	Abg22388 Novel hum
8	898	30.0	167	4 ABG22387	Abg22387 Novel hum
9	869	29.1	181	3 AAB43724	Aab43724 Human can
10	494	16.5	108	5 ABP05496	Abp05496 Human ORF
11	329.5	11.0	2016	4 ABB71487	Abb71487 Drosophil
12	232	7.8	1539	4 AAM78908	Aam78908 Human pro
13	232	7.8	1728	4 AAM39234	Aam39234 Human pol
14	232	7.8	1766	4 AAM79892	Aam79892 Human pro
15	232	7.8	1766	4 AAM41020	Aam41020 Human pol
16	232	7.8	2039	6 ABR64258	Abr64258 Angiogene
17	232	7.8	2039	7 ADD45832	Add45832 Human pro
18	217	7.3	663	4 ABB70523	Abb70523 Drosophil
19	212.5	7.1	386	4 ABB70522	Abb70522 Drosophil
20	203.5	6.8	656	4 ABUS3239	Abus3239 Human tes
21	194	6.5	3238	4 ABB71715	Abb71715 Drosophil
22	185	6.2	636	4 ABG03757	Abg03757 Novel hum
23	181.5	6.1	268	4 AABU16305	Aabu16305 Human nov
24	181.5	6.1	268	6 ABUS5374	Abus5374 Human nov
25	181	6.1	2759	6 AAO16418	Aao16418 Human nuc

26	174.5	5.8	669	2 AAW37483	Aaw37483 Mouse liv
27	174.5	5.8	2289	2 AAW14987	Aaw14987 Protein d
28	174	5.8	107	4 ABB16825	Abb16825 Human ner
29	174	5.8	704	7 ADE60099	Ade60099 Rat Prote
30	174	5.8	704	7 ADD47584	Add47584 Rat Prote
31	174	5.8	704	7 ADE60734	Ade60734 Rat Prote
32	174	5.8	704	7 ADE60097	Ade60097 Rat Prote
33	174	5.8	704	7 ADD47582	Add47582 Rat Prote
34	174	5.8	704	7 ADD47665	Add47665 Rat Prote
35	171.5	5.7	699	4 AAM79156	Aam79156 Human pro
36	169.5	5.7	718	4 AAM80140	Aam80140 Human pro
37	169	5.7	676	4 AAG64373	Aag64373 Human hsp
38	169	5.7	676	4 AAM47218	Aam47218 Human NOV
39	169	5.7	676	4 AAE09441	Aae09441 Human SB-
40	169	5.7	676	5 ABB05039	Abb05039 Human NOV
41	169	5.7	717	4 AAE09442	Aae09442 Novel hum
42	168.5	5.6	1783	5 ABB97448	Abb97448 Novel hum
43	167.5	5.6	1782	6 ABR64223	Abr64223 Angiogene
44	167.5	5.6	1795	4 ABB21018	Abb21018 Novel hum
45	167	5.6	667	5 ABB05040	Abb05040 Human NOV

ALIGNMENTS

RESULT 1
AAY67579
ID AAY67579 standard; protein; 562 AA.
AC AAY67579;
XX
XX 19-JUN-2000 (first entry)
XX
XX Human death inducer-obliterat 1 (DIO-1) polypeptide.
XX
XX Death inducer-obliterat 1; DIO-1; cell death; cancer; tumour; human;
XX autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
XX antirheumatic; antiinflammatory; antiproliferative.
XX
XX Homo sapiens.

Key	Location/Qualifiers
FT Region	165..172
FT Region	/note= "NLS sequence"
FT Region	185..193
FT Region	/note= "NLS sequence"
FT Region	271..288
FT Region	/note= "zinc finger motif"
FT Region	293..320
FT Region	/note= "zinc finger motif"

WO2000015787-A1.

23-MAR-2000.

10-SEP-1999; 99WO-GB003019.

10-SEP-1998; 98SE-00003069.

17-SEP-1998; 98US-0100873P.

(CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.

(BANN/) BANNERMAN D G.

Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;

WPI; 2000-271426/23.

N-PSDB; AAZ90578.

New DNA encoding human and murine death inducer-obliterat 1

polypeptides, useful in the treatment of cancer, autoimmune diseases,

diabetes, rheumatoid arthritis, benign tumors, malignant tumors and

hyperproliferative skin disorders.

XX

PS Claim 6; Fig 1C; 27pp; English.

XX The invention provides nucleic acids encoding the human and murine death

CC inducer-obliator 1 (DIO-1) polypeptides. The polypeptides can be

CC expressed by standard recombinant methodology. The DIO-1 polypeptides,

CC agonists and antagonists are used as a medicament for treating diseases

CC characterized by an alteration in cell death or by hyperproliferation,

CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign

CC tumours, malignant tumours or hyperproliferative skin disorders. They are

CC also useful in the treatment of metabolic, proliferative or inflammatory

CC conditions. The present sequence represents the human DIO-1 polypeptide

XX SQ Sequence 562 AA;

Query Match 100.0%; Score 2989; DB 3; Length 562;

Best Local Similarity 100.0%; Pred. No. 2.3e-228;

Matches 562; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIARREGARRSMFVSLDSGPTSCPATDAETASGVSASE 60

Db 1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIARREGARRSMFVSLDSGPTSCPATDAETASGVSASE 60

QY 61 LRSGROPKRTVERVEQFLTIARRRGRSMFVSLDSGPTSCPATDAETASGVSASE 120

Db 61 LRSGROPKRTVERVEQFLTIARRRGRSMFVSLDSGPTSCPATDAETASGVSASE 120

QY 121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNLRLRRKREQPTERP 180

Db 121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNLRLRRKREQPTERP 180

QY 181 LKGIOSRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240

Db 181 LKGIOSRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240

QY 241 KAAQDIKDESPFCDLGRPKPECEGYDPNALYICICRQPHNNRPMICCDRCCEWPHGDCVGIS 300

Db 241 KAAQDIKDESPFCDLGRPKPECEGYDPNALYICICRQPHNNRPMICCDRCCEWPHGDCVGIS 300

QY 301 EARGLLERNGEDYICPNCTIILQVQDTHSETADQOEAKWRPDADGTDCTSIGTIEOKS 360

Db 301 EARGLLERNGEDYICPNCTIILQVQDTHSETADQOEAKWRPDADGTDCTSIGTIEOKS 360

QY 361 SEDQGIKGRIEKAAKPSGKKLKIPOPVIEAPGASKICIGFCGCHVAQPDVYCSNDCILK 420

Db 361 SEDQGIKGRIEKAAKPSGKKLKIPOPVIEAPGASKICIGFCGCHVAQPDVYCSNDCILK 420

QY 421 HAAATWKFLLSSGKEQKPK 480

Db 421 HAAATWKFLLSSGKEQKPK 480

QY 481 KAVVVPARSEALGKEAACESSTPSWASDHNVAVKPEKTAAPSPSLLKCMYHLGVGLLD 540

Db 481 KAVVVPARSEALGKEAACESSTPSWASDHNVAVKPEKTAAPSPSLLKCMYHLGVGLLD 540

QY 541 PGRSPWIAIPWACPGIGVAALC 562

Db 541 PGRSPWIAIPWACPGIGVAALC 562

RESULT 2

AAB93638

ID AAB93638 standard; protein; 562 AA.

XX

AC AAB93638;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:13130.

XX

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

DF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

FA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DE WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PT

PT

PT

PS Claim 8; SEQ ID NO 13130; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of oligonucleotides comprises a 5'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX

SQ Sequence 562 AA;

Query Match 99.7%; Score 2981; DB 4; Length 562;

Best Local Similarity 99.8%; Pred. No. 9.8e-228;

Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIARREGARRSMFVSLDSGPTSCPATDAETASGVSASE 60

Db 1 MDDKGDPSNEEAPKAIKPTSKFRKTWGFRTTIARREGARRSMFVSLDSGPTSCPATDAETASGVSASE 60

QY 61 LRSGROPKRTVERVEQFLTIARRRGRSMFVSLDSGPTSCPATDAETASGVSASE 120

Db 61 LRSGROPKRTVERVEQFLTIARRRGRSMFVSLDSGPTSCPATDAETASGVSASE 120

QY 121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNLRLRRKREQPTERP 180

Db 121 TRSGPQASATAVKERPASSEKVGKGDHDDTSDSDGLTLKELQNLRLRRKREQPTERP 180

QY 181 LKGIOSRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240

Db 181 LKGIOSRLKRRREGPAETVGSASDTVEGVLPKQRPENDQGVVSOAGKDDRESKLEG 240

QY 241 KAAQDIKDESPFCDLGRPKPECEGYDPNALYICICRQPHNNRPMICCDRCCEWPHGDCVGIS 300

XX

Db 241 KAAQDIKDEEPEDLGRPKPECEGYDPNALYCICROPHNNRPMICCDRCBFWHGBCVGIS 300
 Qy 301 EARGLLRENGEDYICPNCITLQVDETHSETADQOEAWRPDADGTDCTSIGTIEOKS 360
 Db 301 EARGLLRENGEDYICPNCITLQVDETHSETADQOEAWRPDADGTDCTSIGTIEOKS 360
 Qy 361 SEDQIGKGRIRKAAKPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 420
 Db 361 SEDQIGKGRIRKAAKPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 420
 Qy 421 HAAATMKFLSSGKEQKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKK 480
 Db 421 HAAATMKFLSSGKEQKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKKPKK 480
 Qy 481 KAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAPSLLYKCMYHLGVGLLD 540
 Db 481 KAVVVPARSEALGKEAACESSTPSWASDHNYNAVKPEKTAAPSLLYKCMYHLGVGLLD 540
 Qy 541 PSRSFWIAIPWACPGGLGVAALC 562
 Db 541 PSRSFWIAIPWACPGGLGVAALC 562
 RESULT 3
 AAM80219
 ID AAM80219 standard; protein; 1191 AA.
 AC AAM80219;
 DT 06-NOV-2001 (first entry)
 XX Human protein SEQ ID NO 3865.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX WO200157190-A2.
 XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 XX 27-APR-2000; 2000US-00560875.
 XX 20-JUN-2000; 2000US-00598075.
 XX 19-JUL-2000; 2000US-00620325.
 XX 01-SEP-2000; 2000US-00654936.
 XX 15-SEP-2000; 2000US-00663561.
 XX 20-OCT-2000; 2000US-00693325.
 XX 30-NOV-2000; 2000US-00728422.
 XX (HYSB-) HYSBQ INC.
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX N-PSDB; AAK53352.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 455-456; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and/or
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX SQ Sequence 1191 AA;
 Query Match 81.3%; Score 2431; DB 4; Length 1191;
 Best Local Similarity 92.8%; Pred. No. 1.2e-183;
 Matches 465; Conservative 0; Mismatches 0; Indels 36; Gaps 1;
 Qy 65 GROPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASBGSVESASSTRSG 124
 Db 22 GROPKRTERVEQFLTIARRGRSRMPVSLSDSGEPTSCPATDAETASBGSVESASSTRSG 81
 Qy 125 POSASTAVKRPASSEKVKGGDDHDDTSDSDSGLTTLKELQNLRRKRQEPTEPLKGI 184
 Db 82 POSASTAVKRPASSEKVKGGDDHDDTSDSDSGLTTLKELQNLRRKRQEPTEPLKGI 141
 Qy 185 QSRLEKRRRGGPAETVGSASDTVEGVLPSKQEPENDQGVVSOAQKDDRESKLEGKAAQ 244
 Db 142 QSRLEKRRRGGPAETVGSASDTVEGVLPSKQEPENDQGVVSOAQKDDRESKLEGKAAQ 201
 Qy 245 DIKDEEPDGLRPKPECEGYDPNALYCICROPHNNRPMICCDRCBFWHGBCVGISRARG 304
 Db 202 DIKDEEPDGLRPKPECEGYDPNALYCICROPHNNRPMICCDRCBFWHGBCVGISRARG 261
 Qy 305 RLLENGEDYICPNCITLQVDETHSETADQOEAWRPDADGTDCTSIGTIEOKSSEDO 364
 Db 262 RLLENGEDYICPNCITLQVDETHSETADQOEAWRPDADGTDCTSIGTIEOKSSEDO 321
 Qy 365 GIKGRIRKAAKPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 388
 Db 322 GIKGRIRKAAKPSGKKLKI FQVTEAPGASKICIGPGCHVAQPSVYCSNDCILK 381
 Qy 389 IEAPGASKICIGPGCHVAQPSVYCSNDCILKHAATMKFLSSGKEQKPKKPKKPKKPE 448
 Db 382 IEAPGASKICIGPGCHVAQPSVYCSNDCILKHAATMKFLSSGKEQKPKKPKKPKKPE 441
 Qy 449 KPSLPKCGAAGIKLSSVHKRPAPKPKETTVKAVVVPARSEALGKEAACESSTPSWASD 508
 Db 442 KPSLPKCGAAGIKLSSVHKRPAPKPKETTVKAVVVPARSEALGKEAACESSTPSWASD 501
 Qy 509 HNYNAVKPEKTAAPSLLYK 529
 Db 502 HNYNAVKPEKTAAPSLLYK 522
 RESULT 4
 AAM79235
 ID AAM79235 standard; protein; 647 AA.
 XX AC AAM79235;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human protein SEQ ID NO 1897.
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX WO200157190-A2.
 XX 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US004098.
XX PR 03-FEB-2000; 2000US-00496914.
XX PR 27-APR-2000; 2000US-00560875.
XX PR 20-JUN-2000; 2000US-00598075.
XX PR 19-JUL-2000; 2000US-00620325.
XX PR 01-SEP-2000; 2000US-00654936.
XX PR 15-SEP-2000; 2000US-00663561.
XX PR 20-OCT-2000; 2000US-00693325.
XX PR 30-NOV-2000; 2000US-00728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX DR WPI: 2001-476283/51.
XX DR N-PSDB; AAK52368.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX PT in diagnosis and gene therapy.
XX PS Claim 20; Page 4293-4294; 6221pp; English.
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
XX CC sequence listing were missing at the time of publication
XX SQ Sequence 647 AA;
Query Match 77.3%; Score 2311; DB 4; Length 647;
Best Local Similarity 92.5%; Pred. No. 1.7e-174;
Matches 441; Conservative 0; Mismatches 0; Indels 36; Gaps 1;
QY 89 MPVSLDSGPTSCPATDAETASGVSASATRSQPSASTAVKRPASSEKVKGGDDH 148
DB 1 MPVSLDSGPTSCPATDAETASGVSASATRSQPSASTAVKRPASSEKVKGGDDH 60
QY 149 DDTSDSDGLTLKELQNLRLRRKREQPTERPLKGIQSLRLKRRREGPAETVGEASDT 208
DB 61 DDTSDSDGLTLKELQNLRLRRKREQPTERPLKGIQSLRLKRRREGPAETVGEASDT 120
QY 209 VEGVLPSKQEPNDQGVSWQAGKDDRESKLEGAADIDKEEPDGLGRPKPECEGYDPNA 268
DB 121 VEGVLPSKQEPNDQGVSWQAGKDDRESKLEGAADIDKEEPDGLGRPKPECEGYDPNA 180
QY 269 LYCICQPHNNRPMICDRCHEWFHGDVCVISEARGLLERNGEDYICPNCITLQVQDET 328
DB 181 LYCICQPHNNRPMICDRCHEWFHGDVCVISEARGLLERNGEDYICPNCITLQVQDET 240
QY 329 HSTADQOEAKWPRGADGTDCTSIGTIEQKSSDQGIKRIEKAANPSKKKLIKLPQ- 387
DB 241 HSETADQOEAKWPRGADGTDCTSIGTIEQKSSDQGIKRIEKAANPSKKKLIKLPQ- 300
QY 388 -----VIEAPGASKICGPGCCHVAQPDVY 412
DB 301 PGVPPTQLFVLQVLEIAVRSISAFYLLHLCISCKVIEAPGASKICGPGCCHVAQPDVY 360
QY 413 CSNDCILKHAATAATKFTLSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 472
DB 361 CSNDCILKHAATAATKFTLSSGKEQKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 420

QY 473 EKKETTIVKAVVVPARSEALGKEAACESSTPSPWASDHNNYNAVKDEKTAAPSPLYK 529
DB 421 EKKETTIVKAVVVPARSEALGKEAACESSTPSPWASDHNNYNAVKDEKTAAPSPLYK 477
RESULT 5
ABG22389
ID ABG22389 standard; protein; 775 AA.
XX AC ABG22389;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #22380.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX FN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR N-PSDB; AAS86576.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 52748; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 775 AA;
Query Match 77.1%; Score 2305.5; DB 4; Length 775;
Best Local Similarity 76.8%; Pred. No. 6e-174;
Matches 468; Conservative 22; Mismatches 52; Indels 67; Gaps 13;
QY 1 MDDKGDGPNNEAPKAIKTSKEFRWGFRTTIAKREGAGDAEDLEPPPPQQLGLS 60

QY 537 GLDPSRSFWIAIPWACPL 556
Db 536 GFGPSHHL-----GGCLGL 550

RESULT 7
ABG22388
ID ABG22388 standard; protein; 330 AA.

AC ABG22388;
DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22379.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS86575.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 52747; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 330 AA;

Query Match 51.9%; Score 1551; DB 4; Length 330;
Best Local Similarity 89.08; Pred. No. 1.5e-114;
Matches 290; Conservative 0; Mismatches 0; Indels 36; Gaps 1;
QY 235 ESKLEGKAAQDIKDEEPDGLGRPKPECEGYDPNALYICRQPHNRFMIOCDCREEMFHG 294

Db 1 ESKLEGKAAQDIKDEEPDGLGRPKPECEGYDPNALYICRQPHNRFMIOCDCREEMFHG 60
QY 295 DCVGISEARGHLLRNGEDYICNCTILOVODTHSETADQOQAKRPGDADGTDCTSIG 354
Db 61 DCVGISEARGHLLRNGEDYICNCTILOVODTHSETADQOQAKRPGDADGTDCTSIG 120
QY 355 TIEQSSDQGIKRIEKAANPSGKKLKIPOP----- 387
Db 121 TIEQSSDQGIKRIEKAANPSGKKLKIPOP----- 180
QY 388 -----VIEARGASKCIGPGCHVAOPDSVYCSNDCILKHAATMKFLSSGKEQPK 438
Db 181 TLLHCISCKVIEAPGASKCIGPGCHVAOPDSVYCSNDCILKHAATMKFLSSGKEQPK 240
QY 439 PKRNMKMKPKPSLPKCAQAGIKISSVHKRPAPEKKEKTTVKKAVVVPARSEALGKEAAC 498
Db 241 PKRNMKMKPKPSLPKCAQAGIKISSVHKRPAPEKKEKTTVKKAVVVPARSEALGKEAAC 300
QY 499 ESSTPSWASDHNYNAVPEKTAAPSP 524
Db 301 ESSTPSWASDHNYNAVPEKTAAPSP 326

RESULT 8
ABG22387
ID ABG22387 standard; protein; 167 AA.

AC ABG22387;
DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #22378.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS86574.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 52746; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The

polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AB00010-AB030377 represent novel human diagnostic
 CC amino acid sequences. The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 167 AA;
 SQ
 Query Match 30.0%; Score 898; DB 4; Length 167;
 Best Local Similarity 100.0%; Pred. No. 4e-63;
 Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 304 GRLLENGEDYICPNCTILQVDETHSETADQOEAKWPGDADGTDCTSIGTIQKSSD 363
 DB 1 GRLLENGEDYICPNCTILQVDETHSETADQOEAKWPGDADGTDCTSIGTIQKSSD 60
 QY 364 QGIGRIEKAANPSGKKLKIPOPVIEAPGASKICIGPGCCCHVAQPDVSVCNDILKHA 423
 DB 61 QGIGRIEKAANPSGKKLKIPOPVIEAPGASKICIGPGCCCHVAQPDVSVCNDILKHA 120
 QY 424 ATMKFLSSGKKQKPKPKKMKMKPKPSLPKCGAAGIKISSVHKRP 470
 DB 121 ATMKFLSSGKKQKPKPKKMKMKPKPSLPKCGAAGIKISSVHKRP 167

RESULT 9
 AAB43724
 ID AAB43724 standard; protein; 181 AA.
 AC AAB43724;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated protein sequence SEQ ID NO:1169.
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
 KW antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 XX
 OS Homo sapiens.
 XX
 FN WO200005350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005882.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587533/55.
 XX
 DR N-PSDB; AAC77933.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer.
 XX
 PS Claim 11; Page 1790-1791; 2352pp; English.
 XX
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given in

AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;
 CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological diseases and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention
 XX
 XX Sequence 181 AA;
 SQ
 Query Match 29.1%; Score 869; DB 3; Length 181;
 Best Local Similarity 99.4%; Pred. No. 8.8e-61;
 Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 254 LGRPKPECEGYDPAALYICICPHNNRPMICCDRCERWFHGDGVCISARGELLERNSED 313
 DB 10 LGRPKPECEGYDPAALYICICPHNNRPMICCDRCERWFHGDGVCISARGELLERNSED 69
 QY 314 YICFNCITLQVDETHSETADQOEAKWPGDADGTDCTSIGTIQKSSDQGIKRIEKA 373
 DB 70 YICFNCITLQVDETHSETADQOEAKWPGDADGTDCTSIGTIQKSSDQGIKRIEKA 129
 QY 374 ANPSGKKLKIPOPVIEAPGASKICIGPGCCCHVAQ 408
 DB 130 ANPSGKKLKIPOPVIEAPGASKICIGPGCCCHVAHP 164
 RESULT 10
 ABP05496
 ID ABP05496 standard; protein; 108 AA.
 XX
 AC ABP05496;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:10974.
 XX
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 FN WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US010836.
 XX
 PR 30-MAY-2000; 2000US-0206132P.
 PR 29-AUG-2000; 2000US-0228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach MD;

XX WPI; 2002-106308/14.
 DR N-PSDB; ABL15590.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX
 PS Disclosure; SEQ ID NO 10974; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frames, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABL15762 to ABL27252 encode the human ORFX
 CC proteins given in ABP00010 to ABL15500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 108 AA;

Query Match 16.5%; Score 494; DB 5; Length 108;
 Best Local Similarity 87.0%; Pred. No. 2.6e-31;
 Matches 94; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 360 SSEDQGIKGRLEKAAKNSGKKKLIQFQVTEAPGASCKIGFGCHVAQPDVSVCSNDIL 419
 DB 1 SGEDHGIGKGRLEKAAKNSGKKKLIQFQVTEAPGASCKIGFGCHVAQPDVSVCSNDIL 60

QY 420 KHAATWKFSSGKEQKPKRKKMKPKPKPSLPKGAQAGIKISSVH 467
 DB 61 KHAATWKFSSGKEQKPKRKKMKPKPKPSLPKGSVQVQIKISSVH 108

RESULT 11
 ABB71487
 ID ABB71487 standard; protein; 2016 AA.
 AC ABB71487;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 41253.
 DS
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 XF
 XX 23-MAR-2000; 2000US-0191637P.
 XX
 PR 11-JUL-2000; 2000US-00614150.
 PR

XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 DR N-PSDB; ABL15590.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 41253; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB157737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2016 AA;

Query Match 11.0%; Score 329.5; DB 4; Length 2016;
 Best Local Similarity 22.1%; Pred. No. 1.3e-16;
 Matches 159; Conservative 73; Mismatches 220; Indels 269; Gaps 29;

QY 8 SNEAPKAIKPTSKFRKTWGRRTTIKREGAGDAEADPLE-----PPFPQQLG--- 58
 DB 639 ADEIPKLVKMLEST-----AADLPPVPVSLPTNIFPAPMGQGVNAA 681

QY 59 -----LSLRSGRQPKRTERVEOFLTARRGRSMPSVLS 94
 DB 682 AADPNINMTANEDEITADFLQVYVGLIEEDKQ-FEAEVYKQVLA-STEPGLTDAIVMP 739

QY 95 DSGEPTSCPATDAETASGVS-ESASETRSGPQASST-----AVKERPASSEKVKGGDDH 148
 DB 740 TSIPEVDVPOAHTNLLNASLTETPAQSMSTSLPIACSTPSSSVAASTPTSAKVVRG--- 795

QY 149 DTSDSDSGDLTKELQNRLEKREOEPTERPLKGIQSLRKKRREE-----GPAETWGS 203
 DB 796 -----YGRVYLPIETAP-----TTRAKRAAOPFPAPGAATSS 830

QY 204 EA-----SDTVGVLPKQEPENDQGVVVSQAGKDDRESKLE 239
 DB 831 DAGNLSFGESSLDASINQPLNTSSLSNDSPQSGPKRPNP-RBPSMAARSTAPRSKLD 889

QY 240 GKAAQD---IKDEFGDLGRPKCEGVDNALYICROPHNNRMTCCDCRCEWFKGDC 296
 DB 890 ASQNNDFDASESQEDDD-----DPNKLWCICROPHNNRMTCCDCLCDWDFHGTG 938

QY 297 VGISSEARGLLERNGRDYICPNCITLOVODETHSETADQOEAKWRPGDADGTDCTSIGTI 356
 DB 939 VGVTKAMGTDNENKIDWKCPKC-----VKRQERSQPRITDMLVTRPTQP 985

QY 357 EOKSESEOGI---KGRIRKAANPSG-KKKLKIPOVIEAP-----GASK--- 396
 DB 986 EORPSETKVLTTTABIYQVVAAPSAPRRTLPVLTUTVASSPMRIEMAKPAKFPPTGAISHQQ 1045

QY 397 -----CIGFG-----CCHVAQFSDSVYCSNDICLXHA----- 422
 DB 1046 QQQLNFRIGSPGKRISETLCVVCKRPASTSVYCGEGICIRKVAQSAIQAHAAKGLPL 1105

QY 423 ---AATMKFLSSGKEQKPKPKEM----- 443
 DB 1106 QNAGAQLNNNSFDKNNKKKJOLFEDVLROADTVSKYVERINVERKSGRIVTGHMPSAH 1165

QY 444 ---EKKPKPKPS---LPKCGAQAQGIKISSVHR---PAPERKETTIVKAVVVPARS---EAL 492

Db 1166 QFRKVLQNPSEVLPSCTVQS-----ADAEKRLKLGAPRAATSTSEPAVLGVAKKPPGP 1221
 QY 493 GKEAACRSSTPSWASDHYNNAVKP-----BKTA-----APSPSLYTKMVLGVGLDPS 542
 Db 1222 AKLSHPQNTTVQASHQIGISSVRPLAKDKKTKTPTVQAPTFN-----RLAAGKPEPV 1274
 QY 543 R 543
 Db 1275 R 1275
 RESULT 12
 AAM78908
 ID AAM78908 standard; protein; 1539 AA.
 XX
 AC AAM78908;
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1570.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 05-FEB-2001; 2001WO-US0040398.
 XX
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 (HYSE-) HYSEQ INC.
 XX
 Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZH;
 Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 WPI; 2001-476283/51.
 DR N-PSDB; AAK52041.
 XX
 Nucleic acids encoding polypeptides with cytokine-like activities, useful
 in diagnosis and gene therapy.
 PT
 Claim 20; Page 3896-3899; 6221pp; English.
 PS
 XX
 The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 Sequence 1539 AA;
 SQ

Query Match 7.8%; Score 232; DB 4; Length 1539;
 Best Local Similarity 21.8%; Pred. NO. 5.1e-09;
 Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;
 QY 92 SLED-SGEPTSCPAATDAETAGSGSVESASETR-SGPQASASTAVKERPASSEKVK-GGDH 148
 Db 95 SLSDKSHAHPGCLKEPHHPAQTVGHVSHSSQKQCHKPQQAAPAMKTNHSHVKELEHPGVH 154
 QY 149 DDTSDSDSLTLKE-----LQNLRRKREQPTERPLKGIOSRLKRRREGPARTVGSE 204
 Db 155 -----FKBEDKLKKEPKNLQPRQRSSKSFSLDEPPLFPIDNIATIRRE-----GS- 202
 QY 205 ASDTVEGVLPKQEPENDQGVVSQAGKDDRESKLEKAAQDIKDEPGDGLGRPKDECEGY 264
 Db 203 -----DHSSSTFSKY-----MTPTSKQC-GP 222
 QY 265 DPNALYICICROPHNRPMICDRCSEWFGDCVGLSEARGRLLRNGBDYICPNTILQV 324
 Db 223 -----CKKPHGNRFVVGCGRCDDNFHGDVGLSLSQAOQMGEDKRYVVCVKCC---A 271
 QY 325 QDETHSETAD-----QQAENRPGDADGDTCTSIGTIEQKSS-----DQIKGRIEKA 373
 Db 272 EEDKTEILDPDTLENQATVEFHSQXT-MECKLGLSKHTTNDKTYIDTIVKHKVIL 330
 QY 374 ANPSGKKKLTIPQVIEAPGASKICIGPGCCHVAOPDSVYCSNDCILKHAATMKVLSGK 433
 Db 331 KRESCEGR-----NSSDCRDNEIKKQWLAPLRKMGQFV 363
 QY 434 EOKPKPKERKMKPKPSLPKQGAQAGIKISSVHKRPAPERKETTIVKAVVVPARSEALG 493
 Db 364 LPRSSSEKSEKIPKESTIVTC---TGEKASKPGTHKQENKKKVKGV-----LN 412
 QY 494 KEAACESTPS 504
 Db 413 VHPAASAKPS 423
 RESULT 13
 AAM39234
 ID AAM39234 standard; protein; 1728 AA.
 XX
 AC AAM39234;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2379.
 XX
 KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PP 26-DEC-2000; 2000WO-US034263.
 XX
 XX 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58390.
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX Example 4; SEQ ID NO 2379; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX Sequence 1728 AA;

Query Match 7.8%; Score 232; DB 4; Length 1728;
 Best Local Similarity 21.8%; Pred. No. 5.9e-09;
 Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;
 QY 92 SLED-SCEPTSCPATDAETASGVSASSTR-SGPOSASTAVKRPASSEKVK-GGDDH 148
 DB 284 SLSKSHAHPCCLKEPHHPAQTGHVSHSQKQCHKPOQAPAMKTNHVKELHPGVH 343
 QY 149 DDTSDSDSGTLKE-----LQNRLLRKRKEQEPTEPLKGIQSLRKRKRREGPAETVGS 204
 DB 344 ----FKEEDKLKAKPKENLQPRQRSSKSFSLDEPLFPDNIATIRRE-----GS- 391
 QY 205 ASDTVEGVLPKQEPENDQGVWSQAGKDRESKLEKGAADIKDRBPDGLGRPKCEGY 264
 DB 392 -----CKPHGNRFVWGCGRCDNFHGDVGLSLSQAGQGEEDKEVYCVKCC--A 411
 QY 265 DPNALYCIQPHNNRPMICDRCEWPHGDCVGISEARGLLERNGEDYICPNTILQV 324
 DB 412 -----CKPHGNRFVWGCGRCDNFHGDVGLSLSQAGQGEEDKEVYCVKCC--A 460
 QY 325 QDETHSETAD-----QGEAKWRPDADGTDCTSIGTIEQKSE-----DQGIKGRIEKA 373
 DB 461 EEDKKTEILDPTLENQATVEFHSQDNT-MECSEKGLSLRHTTNDRTKYIDDTVKHKVKIL 519
 QY 374 ANPSGKKKLLFPQVIEAPGASKICIGPCCHVAQPOSVYCSNDCILKHAATMKPLSGK 433
 DB 520 KRESGEGR-----NSSDCRDNEIKKQLAPLRKQGQV 552
 QY 434 ECKPPEKMKQKDEKSLPKCGAAGIKLSSVHKRPAPKPKTTTKYAVVPARSEALG 493
 DB 553 LPRSSSEKSEKIPKSTTTVTC---TGEKASKPFTHEKQEMKKQKVEKGV-----LN 601
 QY 494 KEACESSTPS 504
 DB 602 VHPAASAKPS 612

RESULT 14

ID AAM79892

AC AAM79892 standard; protein; 1766 AA.

XX AAM79892;

XX 06-NOV-2001 (first entry)
 DE Human protein SEQ ID NO 3538.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 XX WO200157190-A2.
 XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00693561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAKS3025.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 376-377; 6221pp; English.
 XX The invention relates to polynucleotides (AAKS1456-AAKS3435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAKS2581), 2111
 CC (AAKS2582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX Sequence 1766 AA;

Query Match 7.8%; Score 232; DB 4; Length 1766;
 Best Local Similarity 21.8%; Pred. No. 6.1e-09;
 Matches 94; Conservative 63; Mismatches 154; Indels 120; Gaps 17;
 QY 92 SLED-SCEPTSCPATDAETASGVSASSTR-SGPOSASTAVKRPASSEKVK-GGDDH 148
 DB 322 SLSKSHAHPCCLKEPHHPAQTGHVSHSQKQCHKPOQAPAMKTNHVKELHPGVH 381
 QY 149 DDTSDSDSGTLKE-----LQNRLLRKRKEQEPTEPLKGIQSLRKRKRREGPAETVGS 204
 DB 382 ----FKEEDKLKAKPKENLQPRQRSSKSFSLDEPLFPDNIATIRRE-----GS- 429
 QY 205 ASDTVEGVLPKQEPENDQGVWSQAGKDRESKLEKGAADIKDRBPDGLGRPKCEGY 264
 DB 430 -----DHSSSFESKY-----MWTPSKQC-GF 449


```

265 DPNALYCTCRQPHNRRFMI CDRCEEWFGDCVGISSEARGLLERNGEDYI CENCITLQV 324
450 -----CKXPHGRFRFVWGCGRCDDWPHGDCVGLSQAQQMGEEDEXEYVCVKCC--A 498
325 QDETHSETAD-----QOEAKWRPGDADGTDCTSIGTIEOKSSE-----DQGIKGRIBKA 373
499 EBDKXTEILDPTLENGQATVEFHSQDXT-MECEKLGJLSKHTTNDRTKYIDDVTXHKVKIL 557
374 ANPSGKKKIKIPOPVEAPGASKIGIPCCCHVAQPDSVYCSNDCILKHAATAWKFLSSGK 433
558 KRBSGEGR-----NSSDCRDNEIKKKWLQALPIRKMGQPV 590
434 EOKPXPKEKMKMKPKPSLPKCAQAGIKLSSVHKRPAPKPKETTVCKAAVWVPARSEALG 493
591 LPRSSSEKSEKIPKSESTTVTC---TGSEKSKPCTHEKQEMKKKKVEKGV-----LN 639
494 KEAACESSTPS 504
640 VHPAASAKPS 650

RESULT 15
AAM41020
ID AAM41020 standard, protein; 1766 AA.
XX
AC AAM41020;
XX
XX
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 5951.
XX
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX OS
XX WO200153312-A1.
XX
XX
XX 26-JUL-2001.
XX
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471825.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00652191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX

```

	in gene therapy. A composition containing a polypeptide or polynucleotide	
	of the invention may be used to treat diseases of the peripheral nervous	
	system, such as peripheral nervous injuries, peripheral neuropathy and	
	localised neuropathies and central nervous system diseases, such as	
	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
	utilisation of the activities such as: Immune system suppression,	
	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
	assays for receptor activity, arthritis and inflammation, leukaemia and	
	C.N.S disorders. Note: The sequence data for this patent did not form	
	part of the printed specification	
CC	CC	
XX	CC	
SQ	Sequence 1766 AA;	
	Query Match	7.8%; Score 232; DB 4; Length 1766;
	Best Local Similarity	21.8%; Pred No. 6.1e-09;
	Matches 94; Conservative	63; Mismatches 154; Indels 120; Gaps 1766;
QY	92 SLED-SGEPTSCPNADARTASRGVSASSTR-SGPOSASTAVKRPASSEKVKGDDH	148
Db	322 SLDSKAHPGCLKEPHHPAQTGEVSHSSQCKHEPQQAPAMKTNSHVKELEHPGVGH	381
QY	149 DTDSDSDGLTLKE-----LQNELRRKEQEPTPLKGTQSLRKKRREGPAETVGSE	204
Db	382 ----FKEEDKLAKKPEKNLQPARRSSKFSLDDEPLPIPDNIATIRE-----GS-	429
QY	205 ASDTVGVLPKQEPENDQGVWSAQGDRESKLEGAAQIDKEPGDLGRPKCEGY	264
Db	430 -----DHSSSFESKY-----MTWPSKQC-GF	449
QY	265 DPNALYCICRPHNNRMVICCDRCHEWFHGDCVGISGARGLLERNGEDIYCPNCTILQV	324
Db	450 -----CKKEHNRMVWCGRCDWFHGDVCVLGSLSAQAQMGEBDEKYCVKCC--A	498
QY	325 QDETSETAD-----QOEAKPWPCHADGTDCTSIGTIEQKSE-----DQGIKGRIEKA	373
Db	499 EEDKKTEILDPTLENQATVFEHSBGDT-WECEKLSLKHTINDRTFYIDDTVRKVKVIL	557
QY	374 ANPFGKKKLKIFQVIPBAPGASKCIQPCGCCHAQPSVTCSNDCILKHAATAWKFLSSGK	433
Db	558 KRSGEGEGR-----NSSDCRDNEIKKWQLALPRMKGQVP	590
QY	434 EQKPKPKKKMKMPKSPKPGCAQAGIKLSSVHKTPAPEKETTKVAIVVPARSEALG	493
Db	591 LPRSRSREKKEKPKGSTTTVTC---TGEEKSKPGTGEKEMKKKKVEKGV-----LN	639
QY	494 KEAACESTPS	504
Db	640 VHPAASAKDS	650

Search completed: April 30, 2004, 08:53:58
Job time : 63 secs